Robotics and Autonomous Systems 98 (2017) 42–55

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot

Posture self-stabilizer of a biped robot based on training platform and reinforcement learning

Weiguo Wu *, Liyang Gao [1]

School of Mechatronics Engineering, Harbin Institute of Technology, China

HIGHLIGHTS

- The active training concept is proposed by applying a training platform.
- The training platform disturbs robots on it with amplitude-limited random motions.
- An automatic abstraction method is proposed for the high-dimensional state space.
- A learning posture stabilizer with hierarchical structure is designed.

ARTICLE INFO

Article history:
Received 1 March 2017
Received in revised form 29 June 2017
Accepted 1 September 2017
Available online 18 September 2017

Keywords:
Learning and adaptive systems
Legged robots
Evolutionary robotics
Stability training
State space automatic abstraction
Self-stabilizer

ABSTRACT

In order to solve the problem of stability control for biped robots, the concept of stability training is proposed by using a training platform to exert random disturbance with amplitude limitation on robots that are to be trained. In this work, an approach to achieve a posture stabilizing capability based on stability training and reinforcement learning is explored and verified by simulations. An automatic abstraction method for state space is proposed by using the Gauss basis function and inner evaluation indexes to speed up the learning process. Hierarchical structure stabilizer using the Monte Carlo method is designed according to the concept of variable ZMP. Training samples are extracted from the state transition of the stability training process using balance controllers based on the robot dynamic model. The stabilizers are trained with and without applying the automatic abstraction of state space. Then simulation tests of them are conducted under conditions where the training platform exerts amplitude-limited random disturbances on the robot. Also, the influence of the model errors is studied by introducing deviations of the CoM position during the simulation tests. By comparing the simulation results of two learning stabilizers and the model-based balance controller, it is demonstrated that the designed stabilizer can achieve approximate success rate of the ideal model-based balance controller and exert all the driving ability of the robot under the large disturbance condition of $\pm 30^{\circ}$ inclination of the platform. Also, the effects of the model error can be overcome by retraining using state transition data with the model error.

1. Introduction

It is a prerequisite for various tasks that a biped robot be able to stabilize itself in the actual work environment. Robot samples of stable motion can be generated by the motion planning method considering the dynamics model of the robot. However, the uncertainties of the environment, such as on uneven or sloping ground and with the unknown external forces acting on the robot, will disturb the planned movement. These disturbances can cause the system state to deviate from the planned trajectory. If the deviations have accumulated to a certain extent, the system will become unstable. So the robustness to disturbances in the environment can be seen as an important evaluation index of the self-stabilizing ability of the robot. In this paper, we will study how to get the biped robot to have a strong robustness for coping with environment disturbances within the range of the robot's driving ability.

In our everyday lives, we can observe how both humans and animals in childhood often fall down; it can be ascribed to their weak stabilizing ability. But adults can keep stable and maintain their balance in the environment despite strong disturbances, such as on sloping or uneven ground, bumpy buses and etc. Also they can keep stable even in the case where the dynamics model has changed, such as by carrying heavy objects. This suggests to us that the self-stabilizing ability can be enhanced with the continuous learning

* Correspondence to: 424 Mailbox, Harbin Institute of Technology, 92 West Dazhi Street , Nan Gang District, Harbin, Heilongjiang Province, 150001, China

E-mail addresses: wuwg@hit.edu.cn (W. Wu), gaoly2010@126.com (L. Gao).

[1] Room 1046, Jixie Building, Harbin Institute of Technology, 92 West Dazhi Street , Nan Gang District, Harbin, Heilongjiang Province, 150001, China,.

http://dx.doi.org/10.1016/j.robot.2017.09.001
0921-8890/

in the real environment. In some relevant researches, the stability enhancement after trainings of humans [1,2] and animals [3] also confirmed the above view. But, unlike living organisms, the robots do not have the same biological self-awareness for self-protection. Therefore, in order to get the stabilizing strategy, if the robots directly explore the real environment by random actions, unavoidable damage will occur.

After considering the acquisition of the biological self-stabilizing ability and comparing it with the characteristics of robots, we [4] proposed the concept of stability training for the robots with a 6-DoF series–parallel mechanism training platform (shown in Fig. 1(a)). The mechanism of the designed stability training platform is shown in Fig. 1(b). Then we further proposed the basic idea of the acquisition of a self-stabilizing ability by means of a combination of stability training and learning algorithms; as shown in Fig. 1(c).

In Fig. 1, the stability training system uses the training platform to imitate disturbances in the environment. Through random motions with limited amplitude, the training platform can randomly exert inertial force disturbance on the robot and randomly change the robot's posture relative to the gravity. Furthermore, disturbances caused by uneven terrain can be imitated by fixing a layer with an uneven surface on the moving platform. These features cover the main effects of the external disturbances in the real environment. So, in this way, the disturbances in the real environment can be simulated comprehensively, which ensures that the learning stabilizer of the robot can develop strong robustness to the environment disturbances. During the proposed stability training process, the system state is integrated from the feedback data of the robot sensor system, including the joint position sensors, the gyroscope, contact force sensor on the robot's feet. The system action can be the motion adjustment given by some balance controllers based on the robot's dynamics model. Combining the system states and actions of the robot on the training platform, we have the system state transition, from which the training data of the learning stabilizer can be extracted.

It can be noticed that, in Fig. 1(c), the stability acquisition of the robot is completed in the stability training system, which is separated from applying the trained stabilizer to adjust the robot motion controller. Compared with training in the ordinary lab environment, training on the platform enriches the disturbances exerted on the robot; and compared with training directly in uncertain environments, training on the platform has the advantage that it can easily ensure the safety of the robot. So, in other words, training on the platform makes it more effective to explore the balance control experience for the robot in a broad range of the system state space. This is the main advantage of the proposed method. As a first step into the research, this paper will primarily focus on the learning stabilizer and the training process of it; however the matter of using the trained stabilizer in uncertain environments will not be addressed in this paper.

In addition, considering that the state space of the biped robot is a continuous variable space with high dimensions, the number of state cells needs to be limited in order to enhance learning convergence. Therefore, an automatic abstraction method of the state space based on internal evaluation indexes is proposed in this paper. This method uses Gauss basis function to segment the system state space. By clustering the training data according to the action difference, the regions in the system state space that correspond to the same action are merged.

The training done under 6-DoF complex disturbance requires a great number of searches. Therefore, in this paper, the stability training on the platform will be simplified into the sagittal plane in order to explore a basic case for the proposed idea. The biped robot in the standing posture is simplified to an inverted pendulum. The amplitude-limited random swinging is applied to the robot as disturbances in the sagittal plane.

This paper is organized as follows: Section 2 reviews the related works and highlights the differences brought on in this paper; Section 3 presents the simplified dynamics model of the system consisting of the training platform and the robot; Section 4 presents the proposed automatic abstraction method of state space based on internal evaluation indexes; Section 5 designs the reinforcement learning stabilizer with the Hierarchical structure; Section 6 presents a variable-ZMP balance controller for training data generation; Section 7 carries out the automatic abstraction calculation of the state space and analyzes the abstraction results; Section 8 conducts simulation verifications of the stabilizers whose state spaces are automatically abstracted or uniformly gridded, and the control effect of the two stabilizers is compared with the ideal balance controller based on the precise dynamics model, and the influence of model error on the stability control effect is also studied; Section 9 presents conclusions.

2. Related works

2.1. The balanced control methods based on the dynamics model

The dynamics model based balance control methods calculate the required adjustment of the robot motion according to deviation of the ZMP or system angular momentum. The commonly used control strategies are: the contact force reflex control, the body attitude control, the step adjustment and the angular momentum control. Most of the current biped robot's balance controllers applied one or several of the above strategies, such as Honda's P-series robots [5] and ASIMO [6], HRP series robots [7], Petman [8], Nao [9] and etc. Besides, in order to obtain the stability of certain motions under certain environment disturbances, some special balance controllers were designed. Combining contact force feedback and inertial measurement unit (IMU) feedback, Kim et al. [10] proposed a reflex control method for inclined ground. Seo et al. [11] proposed a balance control method based on torso attitude feedback and virtual gravity compensation, and realized stable walking on uneven ground and inclined ground in experiments. Shahbazi et al. [12] proposed a predictor of the ground inclination based on robot dynamics model and joint angle feedback. Using this predictor, a balance controller was designed to keep the robot standing on a slope with varying inclination. According to the response of the human body to lateral force disturbance, Yoshida et al. [13] proposed a balance control method by combining two control strategies of torso translation and lateral leg lifting. This control method could keep the robot standing stable under the disturbances caused by external impact force or continuous force.

These balance control methods predetermine the adjustment movement according to the robot motion and the disturbances in the environment. For example, the balance control of the standing posture adjusts the movement of the ankle joint or hip joint; balance control during walking adjusts leg movement and step position; when the robot is disturbed by external force, compliant control (making the robot move along the direction of the force) should be conducted first, and then the recovery movement is carried out. Therefore, the versatility of the model-based balance control methods to the robot motion and the disturbances not considered in the design stage is limited. Also, the control effect of such methods depends on the accuracy of the system dynamics model. The biped robots have uncertainties in joint friction, contact with the ground and in many other respects. So it is very difficult to establish an accurate robot-environment dynamics model and obtain accurate model parameters. On account of the above mentioned reasons, most of the dynamics model based balance control methods are only effective in the laboratory-defined, structured environments.