Zbornik radova 49. Konferencije za ETRAN, Budva, 5-10. juna 2005, tom IV
Proc. 49th ETRAN Conference, Budva, June 5-10, 2005, Vol. IV

CONTROL ALGORITHM FOR HUMANOID ROBOTS WALKING BASED ON LEARNING STRUCTURES

Dusko Katic and Aleksandar Rodic
Robotics Laboratory,Mihailo Pupin Institute
Vogina 15, 11060 Belgrade, Serbia & Montenegro
e-mail:dusko,vuk@robot.imp.bg.ac.yu

***Abstract*— In this paper integrated dynamic control of humanoid locomotion mechanisms based on the spatial dynamic model of humanoid mechanism is concerned. The control scheme was synthesized using the centralized model with proposed structure of dynamic controller that involves two feedback loops: position-velocity feedback of the robotic mechanism joints and reinforcement learning feedback around Zero-Moment Point. The proposed reinforcement learning is based on modified version of GARIC architecture for dynamic reactive compensation. Simulation experiments were carried out in order to validate the proposed control approach.**

***Index Terms*— Humanoid robots, Biped locomotion, Dynamically balanced gait, Reinforcement learning.**

I. INTRODUCTION

Many aspects of modern life involve the use of intelligent machines capable of operating under dynamic interaction with their environment. In view of this, the field of biped locomotion is of special interest when human-like robots are concerned. Although there has been a large number of the control methods used to solve the problem of humanoid robot walking , it is difficult to detect a specific trend. Classical robotics and also the more recent wave of humanoid and service robots still rely heavily on teleoperation or fixed behavior-based control with very little autonomous ability to react to the environment. Among the key missing elements is the ability to create control systems that can deal with a large movement repertoire. variable speeds. constraints and most importantly, uncertainty in the real-world environment in a fast, reactive manner. We can however detect a major breakthrough that is definitely setting a new control direction based on introduction of learning and appropriate soft-computing paradigms to biped locomotion . These methods has shown better results in more cases than conventional control methods.

In this paper, a novel, integrated hybrid dynamic control structure for the humanoid robots is proposed, using the complete model of robot mechanism. Our approach consists in departing from complete conventional control techniques by using hybrid control strategy based on model-based approach and learning by experience and creating the appropriate adaptive control systems. Hence, the first part of control algorithm represents some kind of computed torque control method as basic dynamic control method, while the second part of algorithm is modified GARIC reinforcement learning architecture for dynamic compensation of ZMP (Zero-Moment-Point) error.

The reinforcement learning method [4]. [3], [2] cite proposed in this paper is based on the Actor-Critic architecture. The Actor network can be thought of as the control agent, because it implements a policy. The Critic network implements the reinforcement learning part of the control system as it provides policy evaluation and can be used to perform policy improvement. This learning agent architecture has the advantage of implementing both a reinforcement learning mechanism as well as a control mechanism. For the Actor, we selected the two-layer, feedforward neural network with sigmoid hidden units and linear output units. For the Critic, neuro-fuzzy network is proposed. The critic is trained to produce the expected sum of future reinforcement that will be observed given the current values of deviation of dynamic reactions and action.

II. MODEL OF THE SYSTEM

A. Model of the robot's mechanism

The considered mechanism has in total *n=20* DOFs of motion.

Bearing in mind the selected basic link of the mechanism, recursive numerical relations are formed that successively determine angular and translational velocities and accelerations of particular links of the robotic mechanism. Taking into account the dynamic coupling between particular parts (branches) of the mechanism chain one can derive the relation that describes the overall dynamic model of the locomotion mechanism in a vector form [1]:

$$P = H(q) + h(q, \dot{q}) + J^T(q)F \qquad (1)$$

where: $P \in R^{n\times 1}$ is the vector of driving moments at the humanoid robot joints; $F \in R^{6\times 1}$ is the vector of external forces and moments acting at the particular points of the mechanism; $H \in R^{n\times n}$ is the square matrix that describes 'full' inertia matrix of the mechanism shown in Fig. 1; $h \in R^{n\times 1}$ is the vector of gravitational, centrifugal and Coriolis moments acting at *n* mechanism joints; $J(q)$ is the corresponding Jacobian matrix of the system; $n = 20$ is the total number of DOFs (Fig. 1). Special importance in the calculation of the model (1) is

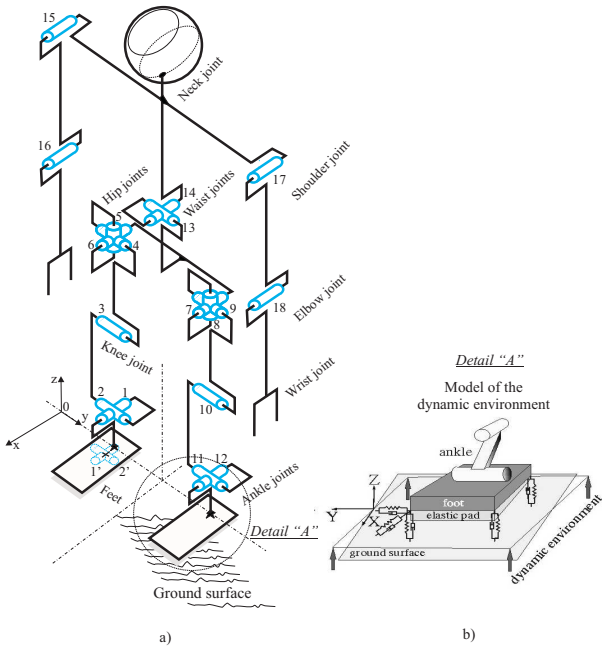

Fig. 1. Model of the humanoid locomotion mechanism with 18 active and 2 passive DOFs: (a) kinematic scheme of the mechanism, (b) dynamic model of the environment

B. Gait phases and indicator of dynamic balance

The robot's bipedal gait consists of several phases that are periodically repeated [1]. Fig. 2 illustrates these gait phases of biped robot locomotion, with the projections of the contours of the right (RF) and left (LF) robot foot on the ground surface, whereby the shaded areas represent the zones of the direct contact with the support. During the walking, the biped is constantly in the state of a certain dynamic balance. The indicator of the degree of dynamic balance is the ZMP, i.e. its relative position with respect to the footprint of the supporting foot of the locomotion mechanism. The Instantaneous position of the ZMP is the best indicator of the dynamic balance of the biped robot.

III. Hybrid Dynamic Integrated Control Algorithm

In accordance with the control task, we propose the application of the algorithm of the so-called hybrid integrated dynamic control, based on the knowing of the overall dynamic model of the system.

In Fig. 3 is presented the block-diagram of the dynamic controller for biped locomotion mechanism, proposed in this work. It involves two feedback loops: (i) position-velocity feedback, (ii) dynamic reaction feedback at the ZMP based on GARIC reinforcement learning structure, The synthesized dynamic controller (Fig. 3) was designed on the basis of the centralized dynamic model.The vector of driving moments $\hat{P}$ represents the sum of the driving moments $\hat{P}_1$ and $\hat{P}_2$. The moments $\hat{P}_1$ are determined so to

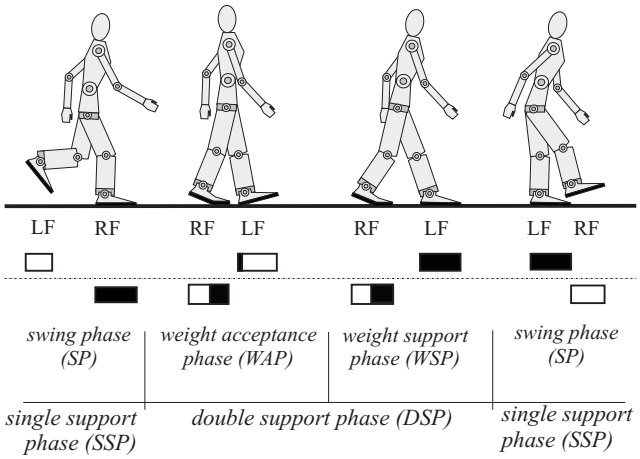

Fig. 2. Phases of biped gait

ensure precise tracking of the robot's position and velocity in the space of joints coordinates.

A. Controller of trajectory tracking

The tracking controller of the locomotion mechanism has to ensure the realization of a desired motion of the humanoid robot and avoiding pf fixed obstacles on its way. In [1], it has been demonstrated how local PD or PID controllers of biped locomotion robots are being designed. In this work, the controller for robotic trajectory tracking was synthesized using the computing torque method in the space of internal coordinates of the mechanism joints:

$$\hat{P} = \hat{H}(q)[\ddot{q}_0 + K_v(\dot{q} - \dot{q}_0) + K_p(q - q_0)] + h(q, \dot{q}) \quad (2)$$

where $\hat{H}$,$\hat{h}$ are the corresponding estimated values of the inertia matrix, vector of gravitational, centrifugal and Coriolis forces and moments from the model (1). The matrices $K_p \in R^{n\times n}$ and $K_v \in R^{n\times n}$ are the corresponding matrices of position and velocity gains of the controller. The gain matrices K_p and K_v can be chosen in the diagonal form by which the system is decoupled into $n$independent subsystems.

B. Reinforcement Learning Compensator of Dynamic Reactions

In the sense of mechanics, locomotion mechanism represents an inverted multi link pendulum. In the presence of elasticity in the system and external environment factors, the mechanism's motion causes dynamic reactions at the robot supporting foot. Thus, the state of dynamic balance of the locomotion mechanism changes accordingly. For this reason it is essential to introduce dynamic reaction feedback at ZMP in the control synthesis. There are relationship between the deviations of ZMP positions ($\Delta x^{(zmp)}$, $\Delta y^{(zmp)}$) from its nominal position 0_{zmp} in the motion directions x and y and the corresponding dynamic reactions $M_x^{(zmp)}$ and $M_y^{(zmp)}$ acting about the mutually orthogonal

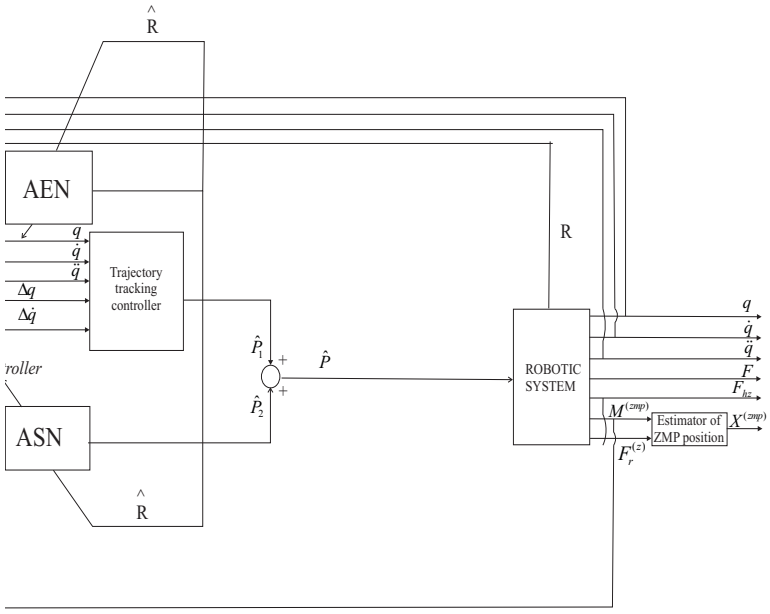

Fig. 3. Block-scheme of the hybrid dynamic control of biped with two feedback loops

axes that pass through the point 0_{zmp}. $M_x^{(zmp)} \in R^{1\times 1}$ and $M_y^{(zmp)} \in R^{1\times 1}$ represent the moments that tend to overturn the robotic mechanism, i.e. to produce its rotation about the mentioned rotation axes (axes of the joints *1'* and *2'* in Fig. 1). On the basis of the above the reinforcement control algorithm [4] is defined with respect to the dynamic reaction of the support at ZMP. In this case external reinforcement signal R is defined according to values of ZMP errors. If ZMP errors in x and y directions are greater then chosen limits of supported polygon, external reinforcement signal is set to value 1. Hence, in this case AEN network (action evaluation network) maps position and velocity tracking errors and external reinforcement signal R in scalar value (internal reinforcement $\hat{R}$) which represent the quality of given control task defined by the chosen control policy:

$$\hat{R}(t+1) = R(t) + \gamma v(t+1) - v(t) \tag{3}$$

where $v(t)$ is output of AEN; γ is a coefficient between 0 and 1. ASN (action selection network) maps the deviation of dynamic reactions in recommended control torque. Exactly, by using SAM(Stochastic action modifier), based on recommended control torque and internal reinforcement $\hat{R}$, control torque P_{dr} is generated. Learning process of AEN (tuning of network weighting factors) is realized by modified version of back propagation algorithm where error is defined by internal reinforcement signal $\hat{R}$. In the same way, using gradient method and internal reinforcement signal, learning process of ASN is realized. $\Delta M^{(zmp)} \in R^{2\times 1}$ is the vector of deviation of the actual dynamic reactions from their nominal values. $P_{dr} \in R^{2\times 1}$ is the vector of control moments at the joints *1'* and *2'* (Fig. 1) that ensures the state of dynamic balance. The control moments P_{dr} calculated from GARIC reinforcement learning structure can not be generated at the joints *1'* and *2'* because these are underactuated, i.e. passive joints. Because of that the control action is 'displaced' to the other, actuated joints of the mechanism chain. Since the vector of deviation of dynamic reactions $\Delta M^{(zmp)}$ has two components about the mutually orthogonal axes *x* and *y,* at least two different active joints have to be used to compensate for these dynamic reactions. Considering the model of locomotion mechanism presented in Fig. 1, the compensation was carried out using the following mechanism joints: *1*, *6* and *14* to compensate for the dynamic reactions about the *x*-axis and *2*, *4* and *13* to compensate for the moments about the *y*-axis. Thus, the ankle joints, hip joints and. waist joints are taken into consideration. Complete control $\hat{P}$ (Fig. 4), is calculated on the basis of the vector of the moments P_{dr} (after distribution it is $\hat{P}_2$ calculated using the GARIC structure , whereby it is borne in mind how many 'compensational joints' are really engaged. In the case when compensation of the ground dynamic reactions is performed using all six proposed joints the compensation moments P_{dr} are uniformly distributed over all of the selected joints, to load uniformly the . In nature, biological systems use simultaneously a large number of joints for correcting their balance. However, for the purpose of verifying the control algorithm, in this work the choice was restricted only to the mentioned six joints: 1, 2, 4, 6, 13 and 14 (Fig. 1).

Beside the proposed control approach, it is important to investigate in future research the following hybrid control

approach:

$$\hat{P} = P_{ff} - K_v(\dot{q} - \dot{q}_0) - K_p(q - q_0)] + P_{rl} \qquad (4)$$

where P_{ff} is feedforward control torque based on centralized nominal dynamic model of biped; P_{rl} - reinforcement learning control structure.

IV. Simulation Experiments

Theoretical results presented previously were analyzed on the basis of numerical data obtained by simulation of the closed-loop model of the locomotion mechanism shown in Fig. 1. The half-step is repeated with this period, whereby the gait phases presented in Fig. 2 alternate regularly. In Figs. 4 and 5 are presented the results of applying the hybrid controller. On analyzing the results presented in Figure. 4 one can see that the we have better results for error of ZMP when algorithm with training of ASN neuro-fuzzy network is used. It can be concluded that without the feedback with respect to the ground reactions around the ZMP it is not generally possible to ensure dynamic balance of the locomotion mechanism in its motion. This comes out from the fact that the nominal trajectory was synthesized without taking into account the possible deviations of the surface on which biped walks from an ideally horizontal plane.

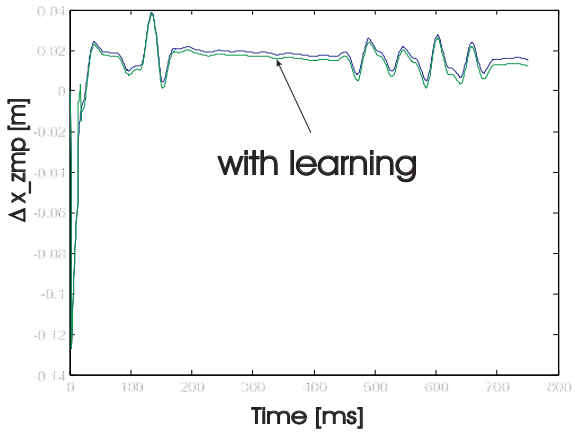

Fig. 4. Error of ZMP in x-direction

In Figure 6 are presented the corresponding deviations (errors) Δq_i and $\Delta \ddot{q}$ of the real values of angles and angular velocities at the robot joints from their nominal values when the controller of tracking desired trajectory was applied. The deviations of the variables converge to a zero value on the given time interval, which means that the controller employed ensured good tracking of the desired trajectory.

In Fig. 6 value of internal reinforcement through process of walking is presented. It is clear that task of walking within desired ZMP tracking error limits is achieved.

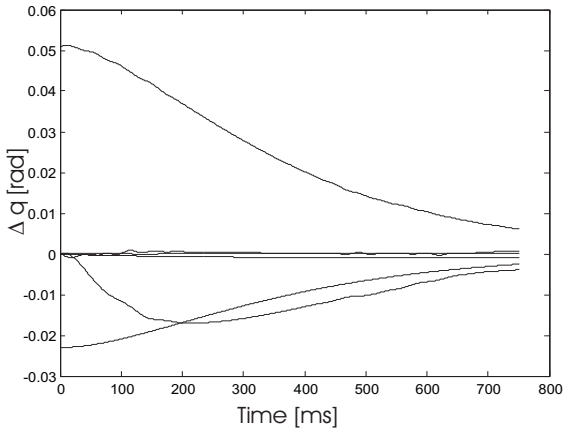

Fig. 5. Position tracking errors for compensation joints

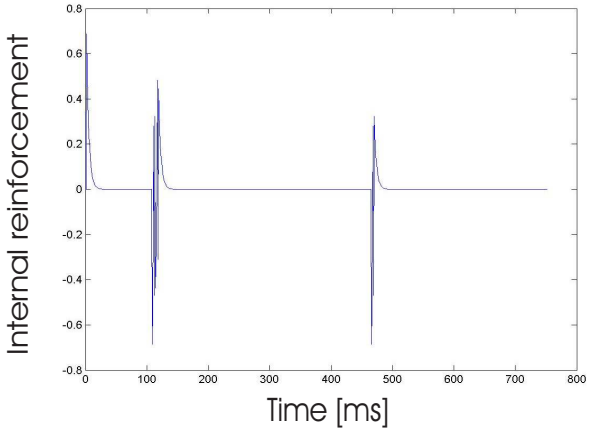

Fig. 6. Internal reinforcement through process of walking

V. Conclusions

Control level consists of the so-called 'basic dynamic controller' was synthesized, consisting of a dynamic controller for tracking robot's nominal trajectory and a compensator of dynamic reactions of the ground around the ZMP based on GARIC reinforcement learning architecture.

REFERENCES

[1] M. Vukobratović, B. Borovac, B. Surla and D. Stokić, *Biped Locomotion: Dynamics, Stability,Control and Application*. Springer-Verlag. Berlin, 1990.

[2] H.Benbrahim and J.A.Franklin",Biped Dynamic Walking using Reinforcement Learning", Robotics and Autonomous Systems", Vol.22, pp.283-302, December 1997.

[3] A.W.Salatian and K.Y.Yi and Y.F.Zheng", "Reinforcement Learning for a Biped Robot to Climb Sloping Surfaces", Journal of Robotic Systems", Vol.14, No.4, pp,283-296, April,199.

[4] H.R.Berenji and P.Khedkar,"Learning and Tuning Fuzzy Logic controllers through Reinforcements", IEEE Transactions on Neural Networks", Vol.3, No.5", pp.724-740, September 1992.