Biped Walk Learning On Nao Through Playback and Real-time Corrective Demonstration

Çetin Meriçli
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213, USA
cetin@cmu.edu

Manuela Veloso
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213, USA
mmv@cs.cmu.edu

Categories and Subject Descriptors

I.2 [**Artificial Intelligence**]: Robotics

General Terms

Algorithms

ABSTRACT

We contribute a two-phase biped walk learning approach which is developed on the Aldebaran Nao humanoid robot. In the first phase, we identify and save a complete walk cycle from the motions of the robot while it is executing a given walk algorithm as a black box. We show how the robot can then play back such a recorded cycle in a loop to obtain a good open-loop walking behavior. In the second phase, we introduce an algorithm to directly modify the recorded walk cycle using real time corrective feedback provided by a human. The algorithm learns joint movement corrections to the open-loop walk based on the corrective feedback as well as the robot's sensory readings while walking autonomously. Compared to the open-loop algorithm and hand-tuned closed-loop walking algorithms, our two-phase method provides an improvement in walking stability, as demonstrated by our experimental results.

Keywords

robot learning from demonstration, skill learning

1. BACKGROUND

Biped walk learning is a challenging problem in humanoid robotics due to the complex dynamics of walking. Developing efficient biped walking methods on commercial humanoid platforms with limited computational power is even more challenging since the developed algorithm should be computationally inexpensive, and it is not possible to alter the hardware.

The Aldebaran humanoid robot, Nao, is a 4.5 kilograms, 58 cm tall humanoid robot with 21 degrees of freedom (http://www.aldebaran-robotics.com/pageProjetsNao.php). Since the introduction of the Nao as the common robot platform of the RoboCup Standard Platform League (SPL) (http://www.tzi.de/spl) in 2008, competing teams have been drawn to investigate efficient biped walking methods suitable for the Nao hardware [1].

Cite as: Biped Walk Learning On Nao Through Playback and Real-time Corrective Demonstration, Author(s), *Proc. of 9th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, van der Hoek, Kaminka, Lespérance, Luck and Sen (eds.), May, 10–14, 2010, Toronto, Canada, pp. XXX-XXX.

There have also been approaches to task and skill learning that utilize learning from demonstration paradigm for biped and quadruped walk learning [2], low level motion planning [3], and task learning for single and multi-robot systems [4].

2. APPROACH

A walking algorithm computes a vector of joint commands and sends it to the robot at each execution timestep. Considering the periodical nature of walking, it is possible to compute a walk cycle once and then play the cycle back in a loop to obtain a continuous walking behavior.

We begin by collecting 10 trials of the robot walking forwards for 3 meters at a fixed speed using Liu and Veloso's walking algorithm [1]. In some trials, the robot completed the walk without losing its balance and in others the robot lost its balance and fell before completing the full walk. A set of walk cycles are extracted from the trials in which the robot completed the walk, yielding a number of repetitions of a complete walk cycle. The extracted walk cycles are then compressed into a single walk cycle by computing the mean value of the joint commands for each individual timestep in the walk cycle. The resulting walk cycle is then played back in a loop to obtain an open-loop walk.

Algorithm 1 Closed-loop walking using sensor-joint couplings.

```
Cycle ← loadWalkCycle()
timestep ← 0
loop
   S_t ← readSensors()
   smoothSensors(Sensors_t)
   for all j ∈ Joints do
      if timestep MOD correctioninterval = 0 then
         C_j = calculateCorrection(Sensors_t, j)
      else
         C_j = 0
      end if
      NextAction_j ← Cycle_j^timestep + C_j
   end for
   timestep ← timestep + 1 (mod CycleSize)
end loop
```

The changes in sensory readings when the robot is about to lose its balance can be used to derive a correction policy by mapping these changes to corrective feedback signals. The correction value for each joint is defined as a function of a single sensor reading. At each timestep the correction values for all joints are computed using the recent sensory readings and the defined correction functions. The pseudo-code of that process is given in Algorithm 1.

We use a human demonstrator to provide corrective feedback signals in real time while the robot is performing playback walk action (Figure 1). One of the major engineering problems in using real time human feedback for bipedal walk learning is that feedback must be provided without interfering with the robot dynamics. We developed a wireless control interface using the Nintendo Wiimote commercial game controller with the Nunchuk extension (http://www.nintendo.com/wii/what/controllers) to provide corrective demonstration to the robot. Both the controller and its extension are equipped with sensors measuring their absolute pitch and roll angles. Rolling a handle causes the corresponding half of the robot to bend towards the direction of the roll (Figure 2).

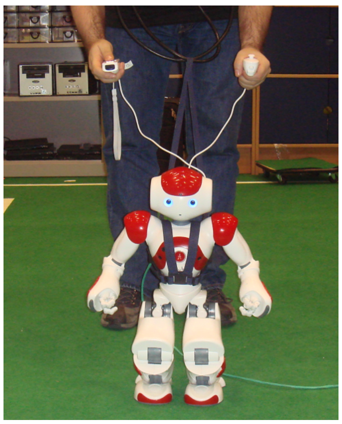

Figure 1: A snapshot from a demonstration session. A loose baby harness is used to prevent possible hardware damage in case of a fall. The harness neither affects the motions of the robot nor holds it as long as the robot is in an upright position.

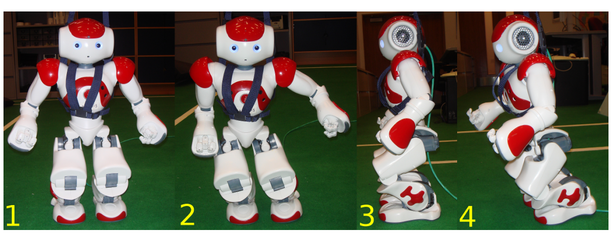

Figure 2: Example corrections using the Wiimote. (a) Neutral position, (b) bent towards right, (c) neutral position, and (d) bent towards forward.

We use the accelerometer readings as the sensory input. We fit a normal distribution on the correction data points received for all 256 possible values of the accelerometer, and we use the means of the normal distributions to compute the correction values during policy execution.

3. RESULTS AND CONCLUSIONS

We evaluated the efficiency of the learned feedback policy against the open-loop policy obtained in the first phase and a hand-tuned closed-loop policy. The hand-tuned policy couples the roll and pitch anglecs of torso computed by torso angle sensor to roll and pitch corrections on both left and right sides using a linear function $C = AX + B$ where X is the sensory reading, with hand-tuned A and B values. We conducted 20 runs per policy and we measured the distance traveled before the robot falls. The results are given in Figure 3.

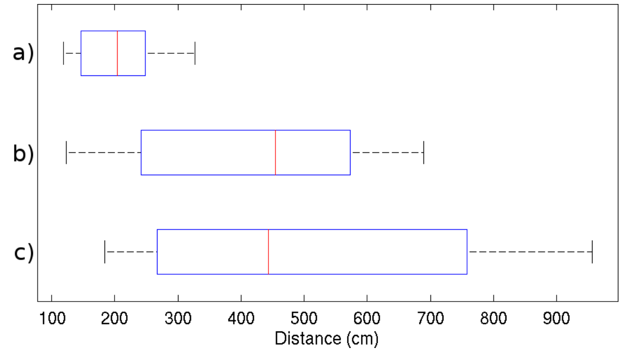

Figure 3: Performance evaluation results: a) open-loop, b) hand-tuned closed-loop using torso angle sensor readings, c) learned policy using accelerometer readings. The lines within the boxes mark the median, the marks at both ends of boxes indicate minimum and maximum distances, and the left and right edges of boxes mark 25^{th} and 75^{th} percentiles, respectively

The learned policy demonstrated an improvement over open-loop and hand-tuned policies, reaching a maximum travelled distance of 956 centimeters while the maximum distances for open-loop and hand-tuned policies were 327 and 689 centimeters, respectively.

This paper contributes a two-phase biped walk learning algorithm and an inexpensive wireless feedback method for providing corrective demonstration. We presented a cheap open-loop playback walking method that learns a complete walking cycle from an existing walking algorithm and then plays it back. This method is highly suitable for humanoid robots with limited on-board processing power.

4. ACKNOWLEDGMENTS

This work is partially supported through The Scientific and Technological Research Council of Turkey (TÜBİTAK) Programme 2214. The authors would like to thank Stephanie Rosenthal, Tekin Meriçli, and Brian Coltin for their valuable feedbacks, and to the members of Cerberus and CMWrEagle RoboCup SPL teams for providing their codebases.

5. REFERENCES

[1] Jinsu Liu, Xiaoping Chen, and Manuela Veloso. Simplified Walking: A New Way to Generate Flexible Biped Patterns. In *Proceedings of the Twelfth International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines (CLAWAR 2009), Istanbul, Turkey, 9-11 September 2009, O. Tosun, H. L. Akin, M. O. Tokhi and G S Virk (Ed.), World Scientific, 2009*, 2009.

[2] D.H. Grollman and O.C. Jenkins. Learning elements of robot soccer from demonstration. In *International Conference on Development and Learning (ICDL 2007)*, pages 276–281, London, England, Jul 2007.

[3] B. Argall, B. Browning, and M. Veloso. Learning robot motion control with demonstration and advice-operators. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'08)*, 2008.

[4] Sonia Chernova and Manuela Veloso. Interactive policy learning through confidence-based autonomy. *Journal of Artificial Intelligence Research*, 34, 2009.