Inverted-Pendulum Mobile Robot Motion Learning from Human Player Observation

Sataya TAKAHASHI*, Hiroaki NONOSHITA, Yasutake TAKAHASHI, and Yoichiro MAEDA
*Dept. of Human & Artificial Intelligent Systems, Faculty of Engineering, University of Fukui
Dept. of Human & Artificial Intelligent Systems, Graduate School of Engineering, University of Fukui
3-9-1, Bunkyo, Fukui, Fukui, 910-8507, Japan
Email: {stakahashi,hnonosita,yasutake,maeda}@ir.his.u-fukui.ac.jp
Takayuki Nakamura
Faculty of Systems Engineering, Wakayama University
Sakaetani 930, Wakayama 640-8510, Japan
Email: ntakayuk@sys.wakayama-u.ac.jp

Abstract—This paper shows an inverted-pendulum mobile robot learns dynamic motion from observation of human player’s demonstration. The robot has torso and body links and learns ball kicking. Before learning, the robot observes human demonstration with a camera, extracts human links on the images, estimates link posture trajectories, and starts kicking motion learning following them. A reasonable and feasible procedure of learning from observation for an inverted-pendulum robot is proposed and investigated in this paper.

I. INTRODUCTION

In order to develop skills, actions, behavior in a human symbiotic environment, a robot is going to learn something from behavior observation of predecessors or humans. Observation of others makes the behavior learning rapid and therefore much more efficient [1], [2], [3]. Recently, robotic imitation methods based on many approaches have been proposed (for example, [4], [5]). It is desirable to acquire various unfamiliar behavior with some instructions from others, for example, surrounding robots and/or humans in real environment because of huge exploration space and enormous learning time to learn. Therefore, behavior learning through observation has been more important.

Many kinds of humanoid robots have been developed so far as human partners under the human-symbiotic environment, because it is reasonable for the mobile robot to have a similar shape and size of a human. However, a two-legged humanoid robot[6] still has a lot of difficulties. For example, it has problems of robustness against unpredictable disturbances, heavy weight, and dynamic motions. A pneumatic actuated biped robot[7][8] has a lightweight body and a potential to generate dynamic motions but still has big problems in terms of the stability for usage under the human living environment.

On the other hand, a two-wheeled inverted pendulum robot[9][10] has many advantages over the statically stable wheeled robots and any other biped robots. It requires a smaller amount of space to stand and stay upright than the other wheeled robots and smaller number of actuators than the conventional biped robots. Furthermore, it can easily generate more dynamic motions only by its body balancing control. The two-wheeled inverted pendulum robot has a possibility to be standard platform as a human partner robot. Most of researches regarding to the two-wheeled inverted pendulum robot focuses on the control theory for achieving stable navigation. There are few researches to realize more dynamic motions such as throwing, jumping, pushing, kicking, and so on.

Takahashi et al.[11] introduced a reinforcement learning method for generating a dynamic motion such that a two-wheeled inverted pendulum robot kicks a ball far away utilizing its own body dynamics while it keeps standing. However, the learning time is too long to apply the learning method for a real mobile inverted-pendulum wheeled robot. Furthermore, they designed initial motions by hand before learning in order to reduce unnecessary useless exploration as much as possible. In order to produce feasible initial motion for efficient learning, imitation is one of the most convincing approaches.

Tamura et al.[12] have proposed a method for recognition of demonstrator’s posture for imitation based on reinforcement learning. In their work, a two-link inverted pendulum wheeled robot learns throwing motion and recognizes posture of demonstrator while the demonstrator shows throwing motion. However, they did not analyze learning performance through the imitation.

In this paper, we investigate learning performance from observation for inverted-pendulum wheeled robot. A human demonstrator shows kicking motion and a robot learns the motion from observed data. A reasonable and feasible procedure of learning from observation for an inverted-pendulum wheeled robot is proposed and investigated.

II. INVERTED PENDULUM WHEELED ROBOT

Fig. 1 shows a model of our inverted pendulum wheeled robot. The robot body size is 20cm length, 30.5cm width, and 70cm height. The weight is 8kg including the torso, body, wheels, batteries, motors, gears, encoders, an accelerometer unit, and control unit. The wheel radius is 8.5 cm and its weight is 500g each. The ball kicked by the robot is a soccer ball, roughly 22cm in diameter, and about 450g
in weight. In order to perform comprehensive experiments for learning a kicking motion with the robot, we develop a computer simulation of the robot and the ball using the ODE(open dynamics engine) library.

III. DESIGN FOR LEARNING KICKING MOTION

The kicking motion is designed with two control layers, a low-level posture controller and a posture generator. The posture controller follows a conventional torque control theory. Torque for the wheels $T$ is calculated as follows:

$$T = -k_1(\theta - \theta_{1d}) - k_2\dot{\theta}_{1d} - k_3\ddot{\theta} - k_4(\varphi - \varphi_d),$$

where $\theta$, $\theta_{1d}$, $\varphi$, $\varphi_d$ are body angle, desired body angle, wheel angular velocity, and desired wheel angular velocity, respectively. $k_1$, $k_2$, $k_3$, $k_4$ are gains for the body angle, body angular velocity, wheel angular velocity, and accumulated error of wheel angular velocity, respectively. For simplicity, torque for a joint between the body and the torso $T_1$ is designed as follows:

$$u_1 = -k_1(\theta - \theta_{1d}),$$

where $k_1$ is a control gain for angular error.

In this work, we design the kicking motion in combination with two primitive motions (Fig.2). Each primitive motion defines the desired body and torso angles, desired wheel angular velocity, gain parameters, and a period of time. The posture generator sends the control parameters to the posture controller within the period defined by the primitive motion, one by one. The first primitive motion tries to lean forward to kick the ball and the second tries to make the robot kick the ball actually. Before and after the kicking motion, the robot controls itself to stay upright avoiding falling-down, autonomously.

IV. POLICY GRADIENT METHOD

In order to generate kicking motion of our robot, our method utilizes a simple policy gradient method introduced by Kohl and Stone[13]. The posture generator learns the control parameters based on the policy gradient method. A parameter set of the current motion is $\Theta$. The learning system prepares $T$ similar policies $R^1, R^2, ..., R^T$ by adding small disturbances $\varepsilon_j$, 0, or $-\varepsilon_j$ to the current motion parameter $\Theta_j(\Theta_j \in \Theta)$:

$$\Theta^r_j = \Theta_j + r\varepsilon_j \quad \text{where} \quad r \in (-1, 0, 1)$$

It evaluates the policies $R^i = \{\Theta^r_j\}$ one by one after the robot tries to generate the kicking motions based on the policies.

First of all, evaluation averages for the disturbances are estimated as follows:

- $\text{Avg}_{+\varepsilon_j}$ is average of evaluation of policies which parameter $\Theta^r_j$ is $\Theta_j + \varepsilon_j$
- $\text{Avg}_{0,j}$ is average of evaluation of policies which parameter $\Theta^r_j$ is $\Theta_j$
- $\text{Avg}_{-\varepsilon_j}$ is average of evaluation of policies which parameter $\Theta^r_j$ is $\Theta_j - \varepsilon_j$

Then, $A_j$, that is, the gradient of evaluation to the policy parameters $\Theta_j$ is estimated approximately as follows: $A_j$ is regarded as 0 if the $\text{Avg}_{0,j}$ is greater than $\text{Avg}_{+\varepsilon,j}$ and $\text{Avg}_{-\varepsilon,j}$. It is regarded as $\text{Avg}_{+\varepsilon,j} - \text{Avg}_{-\varepsilon,j}$, else:

$$A_j = \begin{cases} 0 & \text{if} \quad \text{Avg}_{+0,j} > \text{Avg}_{+\varepsilon,j} \\ & \text{and} \quad \text{Avg}_{+0,j} > \text{Avg}_{-\varepsilon,j} \\ \text{Avg}_{+\varepsilon,j} - \text{Avg}_{-\varepsilon,j} & \text{else} \end{cases}$$

Finally, it updates the policy parameters $\Theta$ according to the $A$ as follows:

$$\Theta_j \leftarrow \Theta_j + \frac{A_j}{|A_j|} \cdot \eta,$$

where $\eta$ is a certain learning step size. The learning system repeats this procedure and updates the motion parameters to reach the local maximum of evaluation.

V. OBSERVATION AND INITIAL LEARNING PARAMETERS

A robot observes a human demonstration of kicking a ball. Fig.3 shows a sequence of observed motion of human kicking behavior. It captures pictures with a camera with 15fps and extracts regions of human on the images with simple background subtraction and frame difference methods. Then, it estimates postures of upper and lower bodies by calculating covariances of the regions on the image. First, it divides the whole human region into two regions, upper body and lower
Fig. 3. View Sequence of Human Kicking Motion

Fig. 4. Upper and Lower Body Extraction from Camera Images of Human Kicking Motion
body, at center of the region. Then, a region moment \( M_{pq} \) is calculated for each region:

\[
M_{pq} = \sum_{i} \sum_{j} (i - i_G)^p (j - j_G)^q f(i, j)
\]

(\( i_G, j_G \)) is the center of the region and \( f(i, j) \) is pixel value on the point \((i, j)\) which is 1 if the point on the region, otherwise zero. The posture \( \theta \) can be calculated as:

\[
\theta = \frac{1}{2} \arctan \left( \frac{2M_{11}}{M_{20} - M_{02}} \right),
\]

where \( p \) and \( q \) are indexes of order of moment for \( x \) and \( y \) coordinates, respectively. Fig.4 shows a sequence of extracted human body regions and estimated posture of upper and lower bodies.

Fig.5 shows a sequence of estimated postures during the observation of human kicking motion.

The parameters were initially set from the observed postures from Fig.5. In general, a human kicking motion does not consist of two primitive motions as described in Section III. A learning robot has to extract useful parameter values from observation of human kicking motion. To be concrete, the observed human motion divided into two primitive motions for generating learning initial parameter values. A sequence of an observed trajectory is divided at points each of which lower body posture acceleration is highest. 0.330 [sec] and 0.594 [sec] are the points to be divided from Fig.5. 0.330 [sec] and 0.264 [sec] are set as periods of time for primitive motion 1 and 2. Posture angles of lower and upper body at 0.330[sec] are -0.302 [rad] and -0.100 [rad], respectively. Then, desired posture angles \( \theta_{1d}, \theta_{2d} \) for primitive motion 1 are set -0.302 [rad] and -0.100 [rad]. In turn, posture angles of lower and upper body at 0.594 [sec] are -0.808 [rad] and 1.30 [rad], respectively. Then, desired posture angles \( \theta_{1d}, \theta_{2d} \) for primitive motion 2 are set -0.808 [rad] and 1.30 [rad]. However, the desired angles from the observation are found as too big to stand stably from a preliminary experiment. Therefore, the desired angle values are set to be 6 times smaller than the values from the observation in this paper. Eventually, the initial parameter values are shown in TABLE I.

![Fig. 5. Posture Estimation during Human Kicking Motion](image)

<table>
<thead>
<tr>
<th>TABLE I</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>INITIAL PARAMETERS OF PRIMITIVE MOTION 1 AND 2</strong></td>
</tr>
<tr>
<td>learning parameter</td>
</tr>
<tr>
<td>period of time [sec]</td>
</tr>
<tr>
<td>( \theta_{1d} ) [rad]</td>
</tr>
<tr>
<td>( \theta_{2d} ) [rad]</td>
</tr>
<tr>
<td>( \phi_{d} ) [rad/sec]</td>
</tr>
<tr>
<td>( k_1 )</td>
</tr>
<tr>
<td>( k_2 )</td>
</tr>
<tr>
<td>( k_3 )</td>
</tr>
<tr>
<td>( k_4 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE II</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>STEP SIZE OF LEARNING PARAMETERS</strong></td>
</tr>
<tr>
<td>learning parameter</td>
</tr>
<tr>
<td>( T ) [sec]</td>
</tr>
<tr>
<td>( \theta_{1d} ) [rad]</td>
</tr>
<tr>
<td>( \theta_{2d} ) [rad]</td>
</tr>
<tr>
<td>( \phi_{d} ) [rad/sec]</td>
</tr>
<tr>
<td>( k_1 )</td>
</tr>
<tr>
<td>( k_2 )</td>
</tr>
<tr>
<td>( k_3 )</td>
</tr>
<tr>
<td>( k_4 )</td>
</tr>
</tbody>
</table>

**VI. KICKING MOTION LEARNING**

**A. Experiment Setup**

As mentioned in section III, each primitive motion is defined by the desired body and torso angles \( \theta_{1d}, \theta_{2d} \), the desired angular linear velocity \( \dot{\theta}_{1d}, \dot{\theta}_{2d} \), the gain parameters \( k_1, k_2, k_3, \) and \( k_4 \), and a period of time \( t \). The learning parameters are those primitive motion parameters for each, therefore, 16 parameters in all. We determine the evaluation criterion of the kicking motion while learning process in consideration of the following points:

- Traveling distance of the robot body should be small.
- Velocity of the kicked ball should be large.
- Robot should avoid falling down by kicking the ball.
- Robot should stand stably after ball kicking.

The actual evaluation is defined as follows:

\[
E = -w_1 l_b + w_2 v_b + w_3 t_f - w_4 \theta_{1f}^2 - w_5 \theta_{2f}^2
\]

where \( l_b, v_b, t_f, \theta_{1f}, \) and \( \theta_{2f} \) are traveling distance of the robot body during kicking motion, the velocity of the kicked ball, time to falling down or end time of trial, body angle and angular velocity at end of trial, respectively. \( w_1, w_2, w_3, w_4, \) and \( w_5 \) are the weights for them and initialized as 1.0, 10.0, 1.0, 1.0, and 1.0, respectively.

The learning step size parameters are shown in TABLE II. The number of deviated policies for estimating gradient of the evaluation \( T \) is set to 50 in the experiments.

**B. Learning Curves**

Fig.6 shows the transition of evaluation and kicked ball velocity while learning kicking motion. Learning curve is like stepwise and somehow the motion evaluation decreases at middle learning stages, for example, from around 1520th to 2350th. The robot learns a ball kicking motion successfully overall although the evaluation suddenly decreases when the
kicked ball speed becomes high because motion tends to be unstable to kick the ball as fast as possible. Fig. 7 shows an example sequence of the learned kicking motion.

**TABLE III**

<table>
<thead>
<tr>
<th>learning parameter</th>
<th>primitive motion 1</th>
<th>primitive motion 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>period of time [sec]</td>
<td>0.500</td>
<td>0.500</td>
</tr>
<tr>
<td>θ₁,θ₂ [rad]</td>
<td>-0.052</td>
<td>0.052</td>
</tr>
<tr>
<td>ϕ₁,ϕ₂ [rad/sec]</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>k₁, k₂</td>
<td>15.0</td>
<td>15.0</td>
</tr>
<tr>
<td>k₃, k₄</td>
<td>0.01</td>
<td>0.01</td>
</tr>
</tbody>
</table>

In order to evaluate the learning curve and its performance, an experiment of learning with hand tuned parameter values. The hand tuned parameters shows in Table III. Fig. 8 shows learning curves. The curves increase stepwise once and go almost flat. Kicked ball velocity and motion evaluation are lower than the ones of which learns with initial parameters of observation shown in Fig. 6.

**VII. CONCLUSION**

In this paper, we investigated learning performance from observation for inverted-pendulum wheeled robot. A human demonstrator shows kicking motion and a robot learns the motion from the observed data. A reasonable and feasible procedure of learning from observation for an inverted-pendulum wheeled robot is proposed. We used the simple policy gradient method to acquire the primitive motions of the kicking motion. Based our approach, we succeeded in acquiring the kicking motion of the two-wheeled inverted pendulum robot and the robot that learned from the observed data kicks a ball stronger than the one from a hand tuned initial parameters. In order to show the proposed method has some generality, we are planning to apply to not only a kicking motion but also, for example, ball throwing, trapping, and so on.

**ACKNOWLEDGMENT**

This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Young Scientists (B), 22700201, 2010.

**REFERENCES**

Fig. 8. Profile of Learning from Hand Tuned Parameter Values


