2013 13th IEEE-RAS International Conference on
Humanoid Robots (Humanoids).
October 15 - 17, 2013. Atlanta, GA

A Humanoid Robot Standing Up Through Learning from Demonstration Using a Multimodal Reward Function

Miguel González-Fierro[1], Carlos Balaguer[1], Nicola Swann[2] and Thrishantha Nanayakkara[3]

Abstract— **Humans are known to manage postural movements in a very elegant manner. In the task of standing up from a chair, a humanoid robot can benefit from the variability of human demonstrations. In this paper we propose a novel method for humanoid robots to imitate a dynamic postural movement demonstrated by humans. Since the kinematics of human participants and the humanoid robot used in this experiment are different, we solve the correspondence problem by making comparisons in a common reward space defined by a multimodal reward function composed of balance and effort terms. We fitted a fully actuated triple inverted pendulum to model both human and robot. We used Differential Evolution to find the optimal articular trajectory that minimizes the Kullback-Leibler difference between the human's and robot's reward profile subject to constraints.**

I. Introduction

Moving from an unstable posture to a stable one, like standing up from a chair, is very often elegantly managed by humans. However, this poses a complex computational problem to humanoid robots that conservatively try to maintain the Zero Moment Point (ZMP) within the support polygon. Learning from demonstration (LfD) is a straightforward way to reconcile this difference. Humans have a high predisposition to learn from demonstration. Some researchers have proposed to denominate our species to be *homo imitans*, which means "man who imitates" [1]. Besides, some researchers defend that LfD is the best way to obtain complex behaviors [2]. Robot LfD follows a set of statements [3]:

LfD1: Determine what to imitate, inferring the goal.
LfD2: Establish a metric for imitation.
LfD3: Mapping between dissimilar bodies.
LfD4: Compute the control commands to perform the imitation.

One of the key questions in LfD is what has been called *what to imitate* or the correspondence problem. It has been demonstrated that when an individual imitates another individual, he does not mimic the same movement or perform the same muscular control orders. On the contrary, he imitates the goal or strategy of the action [4].

Following that idea [4], we propose to use the reward as the goal (LfD1) using a multimodal reward profile, composed of ZMP and torque terms, as the metric of imitation (LfD2). We modelled (LfD3) both human and humanoid as a simple Fully Actuated Triple Inverted Pendulum (FATIP). Finally, we computed an articular trajectory using a PD controller (LfD4), which minimized the Kullback-Liebler difference between human and humanoid reward profiles. We selected Differential Evolution (DE) developed by [5], as the optimizer algorithm.

Different approaches have been presented to address this problem. In [6] a humanoid robot stands up from a chair based on human demonstrations. A three link simulated pendulum that learns to stand up using a hierarchical reinforcement learning method has been presented in [7].

A similar approach to ours, which combines LfD and Reinforcement Learning (RL), teach a robot a pick and place task [8]. The main differences are that they teach the robot by kinesthetic demonstrations which are encoded via Gaussian Mixtures Models (GMM), so they skip the problem of mapping bodies. Also, they use RL to recompute the trajectories if a unplanned obstacle is found.

Many works includes ZMP, torques, joint limits or energy as elements of RL based movements [9], [10]. The main difference with our proposal is the use of the reward as a common basis of comparison between the human and the robot.

Our work is inspired in the work of [11]. The authors define three metrics to solve the correspondence problem to address the problem of mapping actions between the teacher and the learner even if they have different embodiments. One of this metrics, called *trajectory level*, considers the overall goal. In this work, they plan a set of sub-states that has to be reached through optimization, in our case, instead of substates we use a reward profile.

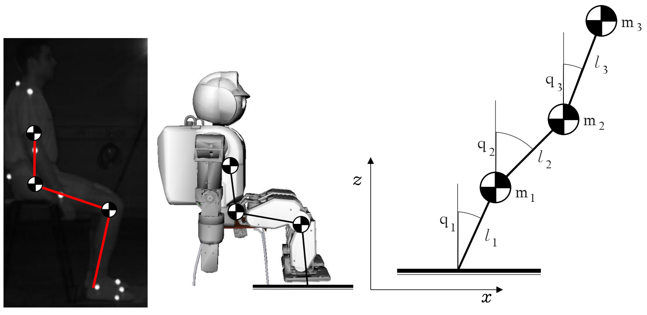
A recent approach [12] address how to obtain a model of the locomotion behavior that can be transferred from a human demonstrator to a robot, what is called *inverse optimal control*. The authors select an objective function which is a combination of position, velocity and other features of the movement as the metrics, multiplied by a set of parameters that are obtained through optimization. This model can be transferred to the robot to produce a similar behavior. The difference with our approach is the selection of the metric to optimize, that in our case is a combination of a reward function of stability and effort. This approach transfers the goal of the movement even though the human and the robot's embodiment are not the same and it can even produce drastically different trajectories, while maintaining the same behavior.

[1] Miguel González-Fierro and Carlos Balaguer are with the Robotics Lab within the Department of System and Automation, Universidad Carlos III, Madrid, Spain mgpalaci@ing.uc3m.es, balaguer@ing.uc3m.es

[2] Nicola Swann is with the School of Life Sciences at Kingston University, Kingston upon Thames, U.K. nicola.swann@kingston.ac.uk

[3] Thrishantha Nanayakkara is with the Center for Robotics Research, Department of Informatics, King's College London, U.K. thrish.antha@kcl.ac.uk

978-1-4799-2619-0/13/$31.00 ©2013 IEEE



(a) (b) (c)

Fig. 1. (a) Snapshot of the high frequency camera of the MOCAP system with a subject seated in a chair and markers in his body. Over him is painted a triple inverted pendulum. (b) A simulation of the humanoid HOAP-3 seated in a chair and the inverted triple pendulum. (c) A triple inverted pendulum in the sagittal plane.

In a previous work [13], we created a method to evaluate and improve the expertise of a group of machine operators. We define a task metric as a combination of several behaviors. The best individual to perform the task is imitated by the other individuals to enhance the task result.

The rest of the paper is organized as follows. Section II explains how the human participant data has been collected and how every human and the robot are modelled as a FATIP. Section III presents the equations of motion and the state space representation of the triple pendulum. In section IV, the proposed algorithm is studied. In section V experimental results are presented and discussed, and finally, the conclusions are stated in section VI.

II. Data collecting and modeling

We collected data from 8 human participants of age between 20 to 40 years, weights between 60 and 99 Kg, and heights between 1.68 and 1.88 m. The experimental protocol was approved by the ethics committee on using human participants in experiments of Kingston University of London. Every participant performed 20 consecutive demonstrations of standing up from a chair. A 6-camera Oqus motion capturing system made by Qualisys, Sweden, collected position data of 21 markers attached to the subject's body at 240Hz sampling rate.

The markers were distributed as follows: first and fifth metatarsi, lateral malleolus (ankle), lateral epicondyle of the femur (knee), greater trochanter (hip), anterior superior iliac spine (ASIS), posterior superior iliac spine (PSIS), seventh cervical vertebra (top of spine), acromion process (shoulder), lateral epicondyle of the humerus (elbow) and lateral styloid process (wrist). All markers are bilateral, they are located on both sides of the body, except the seventh cervical vertebra.

Both robot and human were modeled as a FATIP taking into account only the sagittal plane, since there is no movement in the horizontal or frontal plane when the human stands up. The model selected can not cover multiple contact or floating base effects, however, since standing up is a simple movement, we chose it for simplicity. For a more complete framework for contact modeling in humanoids please refer to [14].

Fig. 1(a) shows a snapshot of the high frequency camera of the MOCAP system, where a human is seated on a chair with all the markers on his body. Over him a triple pendulum is painted. In Fig. 1(b) we show the position of the triple pendulum over the humanoid robot. The center of mass of every pendulum is located at the tip of every link whose masses are negligible. The first joint of the pendulum corresponds to the ankle joint in both human and humanoid, the second joint of the pendulum corresponds to the knee and the third one corresponds to the hip.

To calculate the masses of the pendulum for the human we took into account the total weight of the subject and a estimation of the mean distribution of human body parts presented in [15]. The length of the pendulum links is estimated using the distance between markers. For the first link, the length is the distance between ankle and knee, for the second one, the distance between knee and hip and for the third one, the distance between the hip and the middle of the chest. A pendulum is computed for every subject so we have obtained a total of 8 FATIP and 20 trajectories of standing up for every pendulum.

The robot used for the experiments is the Fujitsu HOAP-3 humanoid. To identify the triple pendulum parameters of the robot, i.e. the length and mass of every link, we used DE and data of the robot sensors like in [16]. We manually created a trajectory for the robot and obtained the ZMP measurement of the FSR sensors in the feet. Later, we used the ZMP multibody equation (1) to obtain the theoretical ZMP trajectory. The multibody ZMP equation in the sagittal plane is

$$x_{ZMP} = \frac{\sum_{i=1}^{n} m_i x_i(\ddot{z}_i + g) - \sum_{i=1}^{n} m_i \ddot{x}_i z_i - \sum_{i=1}^{n} I_{iy}\alpha_{iy}}{\sum_{i=1}^{n} m_i(\ddot{z}_i + g)} \qquad (1)$$

where m_i is the mass of every link, $x_i, z_i, \ddot{x}_i, \ddot{z}_i$ are the position and acceleration of every joint, I_{iy} is the inertia and α_{iy} is the angular acceleration (see Fig. 1(c)).

To identify the system we optimized the pendulum parameters minimizing the quadratic difference between the theoretical ZMP and the real ZMP. The results are shown in the table I.

TABLE I

Triple pendulum identification parameters

	Mass (Kg)	Lenght (m)
Link 1	0.505	0.167
Link 2	0.500	0.260
Link 3	3.900	0.264

III. Fully actuated triple inverted pendulum model

The equation of motion for the FATIP (see Fig. 1(c)) can be obtained using the lagrangian equation:

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}_i}\right) - \frac{\partial \mathcal{L}}{\partial q_i} = \tau_i \qquad (2)$$

where the Lagrangian is the difference between the kinetic and potential energy given by

$$\mathcal{L} = \mathcal{T} - \mathcal{V} \tag{3}$$

$$\mathcal{V} = m_1 g z_1 + m_2 g z_2 + m_3 g z_3 \tag{4}$$

$$\mathcal{T} = \frac{1}{2} m_1 v_1^2 + \frac{1}{2} m_2 v_2^2 + \frac{1}{2} m_3 v_3^2 \tag{5}$$

where v_1, v_2 and v_3 are the speed of the centers of mass of the inverted pendulum and $v_i^2 = \dot{x}_i^2 + \dot{z}_i^2$. Substituting (3), (4) and (5) into (2) we obtain the equation of motion of the triple pendulum, whose compact form is stated as follows

$$\tau = \mathbf{H}(\mathbf{q})\ddot{\mathbf{q}} + \mathbf{C}(\mathbf{q}, \dot{\mathbf{q}})\dot{\mathbf{q}} + \mathbf{G}(\mathbf{q}) \tag{6}$$

where $\mathbf{H} \in \mathbb{R}^{3\times3}$ is the inertia matrix, $\mathbf{C} \in \mathbb{R}^{3\times3}$ is the matrix of centrifugal and coriolis forces and $\mathbf{G} \in \mathbb{R}^{3\times1}$ is the gravity matrix. In the Appendix, the obtantion of (6) is detailed.

A. State space representation of the triple pendulum

The FATIP can be expressed as a dynamical system in the standard form:

$$\dot{\mathbf{x}} = A\mathbf{x} + B\mathbf{u} \tag{7}$$

$$\mathbf{y} = C\mathbf{x} \tag{8}$$

where $\mathbf{x}$ is the state vector, $\mathbf{u}$ is the control vector and $\mathbf{y}$ is the output vector.

To obtain the representation of the triple pendulum system let us define the following state variables: $\mathbf{x} = [q_1, \dot{q}_1, q_2, \dot{q}_2, q_3, \dot{q}_3]^T$.

Taking this into account, and reordering (6), the matrices A, B and C can be obtained given

$$\dot{\mathbf{x}}_\mathbf{1} = \mathbf{x}_\mathbf{2}, \qquad \dot{\mathbf{x}}_\mathbf{3} = \mathbf{x}_\mathbf{4}, \qquad \dot{\mathbf{x}}_\mathbf{5} = \mathbf{x}_\mathbf{6} \tag{9}$$

$$\begin{pmatrix} \dot{\mathbf{x}}_\mathbf{2} \\ \dot{\mathbf{x}}_\mathbf{4} \\ \dot{\mathbf{x}}_\mathbf{6} \end{pmatrix} = \hat{f}(\mathbf{x}_\mathbf{1}, \mathbf{x}_\mathbf{2}, \mathbf{x}_\mathbf{3}, \mathbf{x}_\mathbf{4}, \mathbf{x}_\mathbf{5}, \mathbf{x}_\mathbf{6}) \tag{10}$$

where $\hat{f}$ contains nonlinear terms of the state variables.

To get rid of the nonlinear terms, we linearized over the point of maximum acceleration, $\mathbf{x}_{i0}$, using a Taylor expansion given by

$$\dot{\widetilde{\mathbf{x}}} = \mathbf{A}\widetilde{\mathbf{x}} + \mathbf{B}\widetilde{\mathbf{u}} \tag{11}$$

where

$$A = \frac{\partial f}{\partial \mathbf{x}}\Big|_{\substack{\mathbf{x}=\mathbf{x_0}\\ \mathbf{u}=\mathbf{u_0}}} \quad ; \quad B = \frac{\partial f}{\partial \mathbf{u}}\Big|_{\substack{\mathbf{x}=\mathbf{x_0}\\ \mathbf{u}=\mathbf{u_0}}} \tag{12}$$

and $\widetilde{\mathbf{x}}_\mathbf{i} = \mathbf{x}_\mathbf{i} - \mathbf{x}_\mathbf{i0}$, $\widetilde{\mathbf{u}}_\mathbf{i} = \mathbf{u}_\mathbf{i} - \mathbf{u}_\mathbf{i0}$.

IV. Postural learning from demonstration

To perform the robot standing up and solve the correspondence problem, we cannot compare directly the position or the torques of the human, since in this case the anthropomorphic difference between them is significant. Instead, we define a reward function as a metrics that evaluates the optimality of the overall goal, similarly to the trajectory level of [11].

In this study we chose balance and effort as elements of the reward vector because of the nature of the control task. A robot moving from a seated posture to a stable upright posture, has to transit through meta-stable postures in the sense that the ZMP lies inside of the support polygon most of the time. The robot will not be able to achieve the stable upright posture if it does not bring it to a stand-still posture inside the support polygon before it tips backward. Due to this reason, most conventional ZMP based controllers would generate excessive joint torques to accelerate the body to the upright posture. This can even lead to an overshoot of the ZMP beyond the support polygon leading to tipping forward. To avoid this, the controllers would have to generate excessive counter torques to pull the ZMP back within the support polygon.

A. Postural control and trajectory generation

The desired joint trajectory of the robot is a cubic spline, which is defined as a piecewise polynomial fitted to a set of via points

$$(t_0, q_0^*), (t_1, q_1^*)...(t_k, q_k^*) \tag{13}$$

where $q_i^* \in \mathbb{R}^N$ is the joint via points at time $t_i \in \mathbb{R}$.

Given these via points, there is a cubic trajectory that passes through these points and satisfy a smooth criteria, given by

$$q_i(t) = a_i(t - t_i)^3 + b_i(t - t_i)^2 + c_i(t - t_i) + d_i \tag{14}$$

where a_i, b_i, c_i, d_i are the polynomial coefficients optimized. The complete joint trajectory $q(t) \in \mathbb{R}^N$ is a concatenation of (14) over the time intervals:

$$q(t) = \begin{cases} q_0(t) & \text{if } t_0 \le t < t_1 \\ \vdots & \\ q_k(t) & \text{if } t_k \le t < t_{k+1} \end{cases} \tag{15}$$

We defined a set of two static postures, one with the robot sitting down $q_i(t = 0)$ and another with the robot standing up $q_i(t = t_f)$. The initial and final postures are arbitrary, the initial posture depends on the height of the chair. We could choose different heights with the restriction that the torque do not pass the maximum value allowed for the motors.

The desired joint trajectory is computed as a cubic spline (15) with an initial, middle and final point. The initial and final points correspond to the static postures and the middle point is obtained using DE.

B. Reward function of balance and effort

We defined a reward function for both human and robot and expressed it in terms of balance and effort. The check the balance, we used the ZMP (1) and to check the effort, the torque (6) of the three joints. We selected a gaussian-like function as a base function (16) to evaluate the behavior in the reward space.

$$f(\chi, t) = \exp \frac{-36(\chi(t) - \theta_{med})^2}{2(\theta_{max} - \theta_{min})^2} \tag{16}$$

This function is used to obtain the ZMP reward profile $r_{zmp}(t) = f(ZMP, t)$ and the torque reward profile $r_{\tau_i}(t) = f(\tau_i, t)$. θ_j represents the ZMP minimum, medium and maximum in the case of the ZMP reward function and similar with the torque reward function. The torque is normalized to compute the reward. For simplification, we do not taken into account the contact when the human is seated.

The support polygon for the human is the length of the feet of every subject, that we estimated using the MOCAP system. For the torque, we used (6) to obtain the maximum and minimum torque for the three joints. The support polygon and torque for the robot is obtained using the manual provided by the manufacturer.

The total reward function is the sum of balance and effort functions, which is given by

$$r(t) = w_{zmp}(t) r_{zmp}(t) + w_{\tau}(t) \frac{\sum_{i=1}^{3} r_{\tau_i}(t)}{3} \tag{17}$$

where $w_i(t)$ are weights to modulate the importance of balance vs effort.

We defined the fitness function (18) to minimize, as the summatory in every time step k of the Kullback-Liebler divergence between the mean reward profile of all the human participants $p(i)$ and the reward profile of the robot $q(i)$. Furthermore, we added as a constraints (19) the ZMP limits, torque limits and joint limits.

$$\min\ g = \sum_k \sum_i p(i) log \frac{p(i)}{q(i)} \tag{18}$$

subject to

$$\theta_{min} \leq \theta \leq \theta_{max} \tag{19}$$

where θ represents ZMP, torque or joint position.

Achieving stability without excessively straining the joint actuators becomes an interesting control feature we wish to acquire from human demonstrators. It was interesting to note from our human demonstrations, initially the ZMP stays outside the shape of the feet and moves to the center when the movement is finished. Furthermore, the torques tend to decrease, specially in the second joint which supports the weight of the upper body, until they became minimum in the upright position. The selection of the base function (16) implies the former situation. When the ZMP is in the middle of the feet and when the torques are zero, the reward is maximum, otherwise the reward descends until zero when the ZMP is outside the limits or the torque surpass the allowed maximum torques. Then the reward function (17), acts as an attractor from a initial static posture to the final static posture and makes the movement possible, imitating the human demonstrations, and at the same time, solving the correspondence problem.

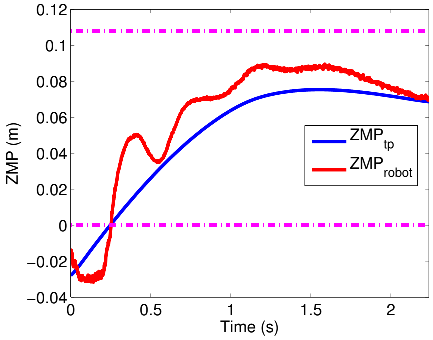


Fig. 2. Computed FATIP ZMP and real robot ZMP. Limits in dotted pink.

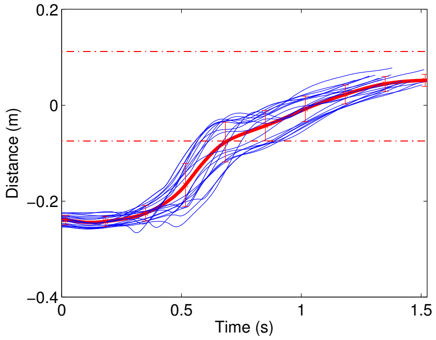


Fig. 3. In blue the ZMP of the 20 trials of a human subject of 67.8Kg and 1.71m. In red the mean ZMP and the standard deviation. In dotted red the mean ZMP limits.

V. Experimental results

The experimental results show that the output trajectory is completely different to that of the human participants, which makes sense. It is obvious that a robot of 60 cm does not stand-up using the same trajectory as a human of 1,80 cm. This result prove our statement and that of [11] and some neuroscientists like [4], that suggest that to imitate a behavior what should be copied is the overall goal.

Fig. 2 shows the theoretical ZMP, calculated using (1) and the real ZMP measured from the robot feet FSR sensors. As it can be seen, initially the ZMP is outside the stability region. This happens because at that time the robot is slightly leaned on the chair and as explained before only contact with the floor is taken into account. In Fig. 3 the ZMP trajectory of one of the human participant is shown. As it can be seen, the ZMP of the human and that of the robot is not the same.

Fig. 4 plots the torques of the robot's pendulum. As it can be seen, they are between the limits. It is remarkable that the second joint has the higher value, this is due it supports the heaviest part of the robot.

In Fig. 5, the calculated rewards for all humans and robot are shown. For every human and every stand up

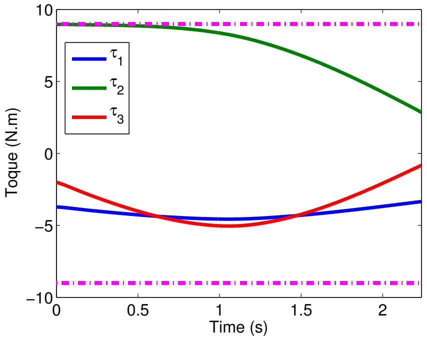


Fig. 4. FATIP joint torques

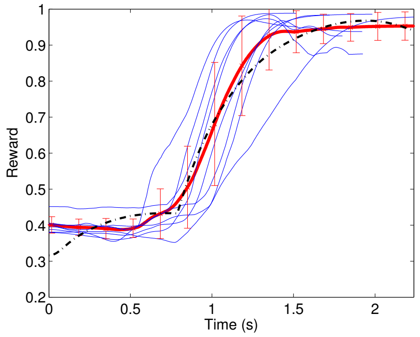


Fig. 5. Mean reward profile for the 8 subjects (in thin blue), mean reward and standard deviation of the humans (in thick red) and optimized robot reward (in dotted black).

demonstration, a reward profile was computed. The mean reward of every human is plotted in blue. The mean of all the 8 mean rewards is plotted in red with the standard deviation. This is the value used in (18) to obtain the desired robot trajectory. Finally, the dotted black line represents the reward profile of the robot.

We implemented our method in the humanoid robot HOAP-3. In Fig. 6 a snapshots of the robot performance are shown. As it can be observed, the robot starts seated in the chair and stands up maintaining the balance in a very soft way. It is clear that the robot does not start with the knee joint at 90 degrees as the human. This is due to the robot's torque limits that are constrained to not be surpassed (see (18) and (19)).

A. Discussion and contributions

The main contribution of this work is the solution of the correspondence problem in the reward space. The transfer of behavior between human and robot is based on the reward obtained when standing up, and is based on the combination of balance and effort. Our method also takes into account the ZMP, torque, and joint limits of the robot, so the trajectory is always executable.

The reward profile is defined as the action goal and may be presented as an evaluation of the behavior success. A successful imitation of the human behavior can be measured taking into account the degree of similarity with the reward profile of the robot.

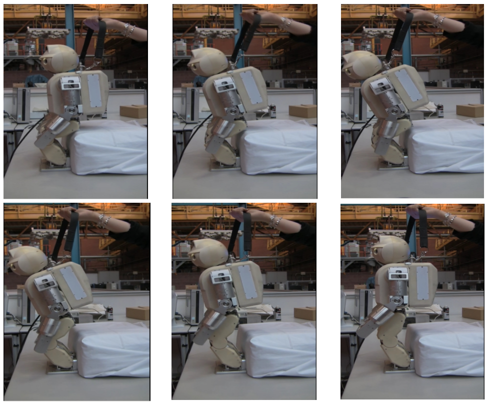


Fig. 6. Snapshots of the experiments of the HOAP-3 humanoid standing up.

The humanoid learns how to perform smooth and stable standing up movements based on human demonstrations, even with a clear mismatch in the embodiment. This is possible because the robot does not simply imitate the human movement, rather learns an optimal behavior subject to a set of internal constraints, which in fact is completely different from the human movement.

Here, we address the correspondence problem by comparing the human demonstrations and robotic behaviors in a reward space defined by a multi-objective reward vector. However, the specific reward functions we have chosen, stability and effort, may neglect other subtle criteria used by human demonstrators. Techniques such as inverse-reinforcement based learning and genetic programming can be used to discover the hidden reward functions in the future. Though such exhaustive exploration for reward functions is beyond the scope of this paper, open exploration for a detailed reward vector will further improve the choices for a robot to innovate diverse skills by selectively focusing on alternative reward functions.

VI. Conclusions and Future Work

In this paper, we proposed a novel method where a humanoid robot learns to stand up from human demonstrations. The human and the robot are very different in terms of height and weight. However, since they are anthropomorphically similar, their behavior must be the same even if they do not perform the same movement. Recent advances in neuroscience [4] suggest that what humans copy when imitating is the overall goal of the behavior, not just the trajectories of the movement.

In that sense, we propose a method where a multi-objective reward function is used to transfer a behavior between a human and a robot. This reward function is the basis of comparison between them. We used Differential Evolution to optimize the desired robot trajectory, which minimizes the Kullback-Liebler difference between the human reward and the robot reward, while taking into account the ZMP, torque and joint limit constraint. We demonstrated that the answer to *what to imitate* question [3] can be to imitate the overall goal of the behavior, defined as a reward profile. The algorithm was tested in the humanoid HOAP-3.

In future works we will address the generalization of our method to other behaviors as opening a door or walking. In these case the reward function has to be selected carefully, or could even be directly learned form the human demonstrations.

Appendix
Fully Actuated Triple Inverted Pendulum equations

Let us define the position and velocity of every link (see Fig. 1(c)).

$$x_1 = l_1 \sin q_1, \qquad \dot{x}_1 = l_1 \cos q_1 \dot{q}_1 \tag{20}$$

$$z_1 = l_1 \cos q_1, \qquad \dot{z}_1 = -l_1 \sin q_1 \dot{q}_1 \tag{21}$$

$$x_2 = l_1 \sin q_1 + l_2 \sin q_2 \tag{22}$$

$$\dot{x}_2 = l_1 \cos q_1 \dot{q}_1 + l_2 \cos q_2 \dot{q}_2 \tag{23}$$

$$z_2 = l_1 \cos q_1 + l_2 \cos q_2 \tag{24}$$

$$\dot{z}_2 = -l_1 \sin q_1 \dot{q}_1 - l_2 \sin q_2 \dot{q}_2 \tag{25}$$

$$x_3 = l_1 \sin q_1 + l_2 \sin q_2 + l_3 \sin q_3 \tag{26}$$

$$\dot{x}_3 = l_1 \cos q_1 \dot{q}_1 + l_2 \cos q_2 \dot{q}_2 + l_3 \cos q_3 \dot{q}_3 \tag{27}$$

$$z_3 = l_1 \cos q_1 + l_2 \cos q_2 + l_3 \cos q_3 \tag{28}$$

$$\dot{z}_3 = -l_1 \sin q_1 \dot{q}_1 - l_2 \sin q_2 \dot{q}_2 - l_3 \sin q_3 \dot{q}_3 \tag{29}$$

The components of every matrix in (6) can be expressed as:

$$\begin{pmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \begin{pmatrix} \ddot{q}_1 \\ \ddot{q}_2 \\ \ddot{q}_3 \end{pmatrix} + \tag{30}$$

$$+ \begin{pmatrix} 0 & c_{12} & c_{13} \\ c_{21} & 0 & c_{23} \\ c_{31} & c_{32} & 0 \end{pmatrix} \begin{pmatrix} \dot{q}_1^2 \\ \dot{q}_2^2 \\ \dot{q}_3^2 \end{pmatrix} + \begin{pmatrix} g_1 \\ g_2 \\ g_3 \end{pmatrix} \tag{31}$$

$$h_{11} = {l_1}^2 \left(m_1 + m_2 + m_3\right) \tag{32}$$

$$h_{22} = {l_2}^2 \left(m_2 + m_3\right) \tag{33}$$

$$h_{33} = {l_3}^2 m_3 \tag{34}$$

$$h_{12} = h_{21} = (m_2 + m_3) l_1 l_2 cos(q_1 - q_2) \tag{35}$$

$$h_{13} = h_{31} = m_3 l_1 l_3 cos(q_1 - q_3) \tag{36}$$

$$h_{23} = h_{32} = m_3 l_2 l_3 cos(q_2 - q_3) \tag{37}$$

$$c_{12} = -c_{21} = -(m_2 + m_3) l_1 l_2 sin(q_2 - q_1) \tag{38}$$

$$c_{13} = -c_{31} = -m_3 l_1 l_3 sin(q_3 - q_1) \tag{39}$$

$$c_{23} = -c_{32} = -m_3 l_2 l_3 sin(q_3 - q_2) \tag{40}$$

$$g_1 = -g l_1 \left(m_1 + m_2 + m_3\right) \sin\left(q_1\right) \tag{41}$$

$$g_2 = -g l_2 \left(m_2 + m_3\right) \sin\left(q_2\right) \tag{42}$$

$$g_3 = -g l_3 m_3 \sin\left(q_3\right) \tag{43}$$

Acknowledgments

The research leading to these results has received funding from the ARCADIA project DPI2010-21047-C02-01, funded by CICYT project grant on behalf of Spanish Ministry of Economy and Competitiveness and from the RoboCity2030-II-CM project (S2009/DPI-1559), funded by Programas de Actividades I+D en la Comunidad de Madrid and cofunded by Structural Funds of the EU.

References

[1] A. Meltzoff, "The human infant as homo imitans," *Social learning: Psychological and biological perspectives*, pp. 319–341, 1988.

[2] B. Argall, S. Chernova, M. Veloso, and B. Browning, "A survey of robot learning from demonstration," *Robotics and Autonomous Systems*, vol. 57, no. 5, pp. 469–483, 2009.

[3] S. Schaal, A. Ijspeert, and A. Billard, "Computational approaches to motor learning by imitation," *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, vol. 358, no. 1431, pp. 537–547, 2003.

[4] G. Metta, G. Sandini, L. Natale, L. Craighero, and L. Fadiga, "Understanding mirror neurons: a bio-robotic approach," *Interaction studies*, vol. 7, no. 2, pp. 197–232, 2006.

[5] R. Storn and K. Price, "Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces," *Journal of global optimization*, vol. 11, no. 4, pp. 341–359, 1997.

[6] M. Mistry, A. Murai, K. Yamane, and J. Hodgins, "Sit-to-stand task on a humanoid robot from human demonstration," in *Humanoid Robots (Humanoids), 2010 10th IEEE-RAS International Conference on*. IEEE, 2010, pp. 218–223.

[7] J. Morimoto and K. Doya, "Acquisition of stand-up behavior by a real robot using hierarchical reinforcement learning," *Robotics and Autonomous Systems*, vol. 36, no. 1, pp. 37–51, 2001.

[8] F. Guenter, M. Hersch, S. Calinon, and A. Billard, "Reinforcement learning for imitating constrained reaching movements," *Advanced Robotics*, vol. 21, no. 13, pp. 1521–1544, 2007.

[9] P. Kormushev, B. Ugurlu, S. Calinon, N. G. Tsagarakis, and D. G. Caldwell, "Bipedal walking energy minimization by reinforcement learning with evolving policy parameterization," in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*. IEEE, 2011, pp. 318–324.

[10] R. Vuga, M. Ogrinc, A. Gams, T. Petric, N. Sugimoto, and A. Ude, "Motion capture and reinforcement learning of dynamically stable humanoid movement primitives," in *Robotics and Automation, 2013. ICRA 2013. IEEE International Conference on*. IEEE, 2011.

[11] A. Alissandrakis, C. Nehaniv, and K. Dautenhahn, "Imitation with alice: Learning to imitate corresponding actions across dissimilar embodiments," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol. 32, no. 4, pp. 482–496, 2002.

[12] K. Mombaur, A. Truong, and J. Laumond, "From human to humanoid locomotionan inverse optimal control approach," *Autonomous robots*, vol. 28, no. 3, pp. 369–383, 2010.

[13] T. Nanayakkara, F. Sahin, and M. Jamshidi, *Intelligent Control Systems with an Introduction to System of Systems Engineering*. CRC Press, 2009.

[14] L. Sentis, J. Park, and O. Khatib, "Compliant control of multicontact and center-of-mass behaviors in humanoid robots," *Robotics, IEEE Transactions on*, vol. 26, no. 3, pp. 483–501, 2010.

[15] NASA, "Man-systems integration standards," NATIONAL AERONAUTICS AND SPACE ADMINISTRATION, Tech. Rep., 1995.

[16] H. Tang, S. Xue, and C. Fan, "Differential evolution strategy for structural system identification," *Computers & Structures*, vol. 86, no. 21-22, pp. 2004–2012, 2008.