Capture, Recognition and Imitation of Anthropomorphic Motion

Sovannara Hak, Nicolas Mansard, Oscar Efrain Ramos Ponce, Layale Saab, Olivier Stasse

2012 IEEE International Conference on Robotics and Automation
RiverCentre, Saint Paul, Minnesota, USA
May 14-18, 2012

Capture, recognition and imitation of anthropomorphic motion

Sovannara Hak, Nicolas Mansard, Oscar Ramos, Layale Saab and Olivier Stasse

I. INTRODUCTION

We present an overview of our current research works in generation, recognition and editing of anthropomorphic motion using a unified framework: the stack of tasks [1]. It is based on the task function formalism classically used for motion generation [2]. A task function maps the joint space of a robot to a dedicated space which is usually linked to the sensors of the robot: the task space. The task spaces are suitable to perform motion analysis and task recognition because the tasks are described in those spaces [3]. The generation is originally based on inverse kinematics but can be generalized to produce full-dynamic motions [4]. The tasks are defined by a tasks space, a reference behavior and a task Jacobian. The reference behaviors are originated from human trajectories. Specific tasks are then integrated to retarget and to edit the reference motion in order to respect the dynamic constraints, the limits of the robot and the general aspect [5]. In the next section, we quickly introduce the stack of tasks framework. Then we present our methods to perform task recognition, dynamic retargeting and editing based on that framework.

II. STACK OF TASKS

The task-function approach [2] consists in designing the motion to be performed as a control law in a subspace of small dimension, and then back-projecting this control law to the state space of the robot. A task is defined by the triple $(\mathbf{e}, \dot{\mathbf{e}}^*, \mathbf{G})$, where $\mathbf{e}$ belongs to the task space, $\dot{\mathbf{e}}^*$ is the reference behavior in the tangent space to the task space at $\mathbf{e}$, and $\mathbf{G}$ is the differential mapping between the task space and the control space of the robot. The interest of defining the robot motion inside a task space rather than directly at the joint level is double: first, the task space is chosen such that the control law can be easily designed (typically, in visual servoing the task space is the space of measurable visual features), making the link between sensor feedback and control direct [6]; second, the interference between two task spaces can be easily prevented and then concurrent simultaneous objectives can be decoupled, using a projection operator. Based on the redundancy of the system, this approach can be extended to consider a hierarchical set of tasks [7]. Complex motion can then be constructed from simple tasks seen as atomic bricks of motion. We define a generic task function formulation [4] that can be applied in both inverse kinematics and inverse dynamics. A complete implementation of this approach is developed in [1] under the name *Stack of Tasks*. The structure enables to easily add or remove a task. The stack of tasks can be generalized to generate dynamic motions.

LAAS-CNRS (Laboratoire d'Analyse et d'Architecture des Systèmes), 7 av col Roche, F-31077 Toulouse, France, Université de Toulouse; UPS, INSA, INP, ISAE: sovannara.hak@laas.fr, nicolas.mansard@laas.fr, oscar.ramos@laas.fr, layale.saab@laas.fr, olivier.stasse@laas.fr.
This work was supported by the R-Blink Project, Contract ANR-08JCJC-0075-01.

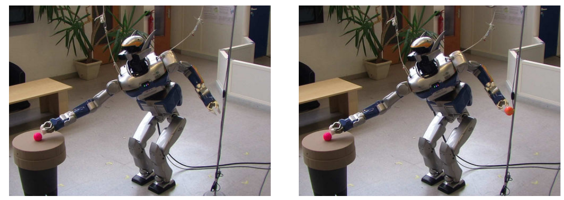

Fig. 1. The final positions of two similar looking movements. Their purposes are different: in the left picture, the robot has to grab one ball, whereas in the right picture, the robot has to grab two balls.

III. TASK RECOGNITION

Several approaches for the representation of an action and its recognition are studied in computer vision, robotics and artificial intelligence [8]. The action recognition and motion analysis is widely handled using statistic tools [9]. The recognition problem is formulated as a classification problem using a Bayesian classifier [10], [11]. Generally, for those statistic based method, the main assumption is that the recognition is performed on a temporal sequence of action. The motion recognition is then divided in two steps, motion segmentation and motion classification [12], [13]. The analysis of the motion has to be performed in a suitable space to be efficient. These spaces can be chosen arbitrary [12], automatically selected [14] or learned [15].

We rely on the task spaces defined using the task function formalism to perform the motion analysis. Assuming that the analyzed motion has been generated by a stack of tasks involving tasks belonging to a known tasks pool, the recognition problem is handled by applying a reverse engineering of the motion. in order to reconstruct the original stack of tasks. The joint trajectory to analyze is projected in a given set of known task spaces. The projected trajectories are compared with theoretical behaviors to decide which tasks are active. Our method is able to recognize tasks executed in parallel and can handle tasks coupling using task spaces and nullspace projectors. Similar looking motion can then be disambiguated. For example, Fig. 1 illustrates the final positions of two different movements played by the HRP-2 robot. Although they look similar, the purpose of the movements are different. In the first movement, the robot has to grab one ball with its right hand. This task modifies the balance of the robot which consequently moves its left

978-1-4673-1404-6/12/$31.00 ©2012 IEEE

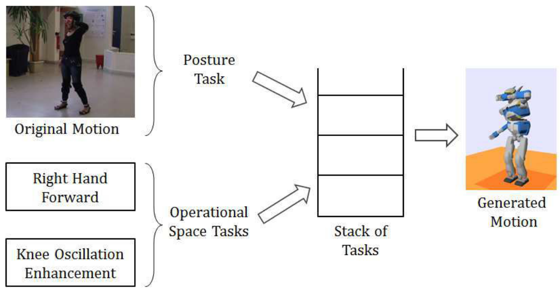

Fig. 2. Inverse Dynamics Cascade Scheme

hand to compensate the balance. In the second movement, the robot has to grab two balls. The second ball is placed at the final position of the left hand in the previous movement. This artificially introduces a visual ambiguity between the two movements. This ambiguity can be solved using our method [3].

IV. Dynamic retargeting and editing

Generating motion from imitation has been adopted widely by researchers in both fields of computer animation and robotics. The starting point is usually the motion acquired from a human expert using a motion capture system [16].

The easiest way to make a humanoid robot behave like a human, is to simply copy human movements. However, the challenges arise due to the kinematic and dynamic disparity between the human and the humanoid. The original movement has to be modified to enforce the kinematic and dynamic constraints of the robot. This operation is called motion retargeting [17], [18]. For example, the retargeting can be done by optimization [19]. Typically, previous attempts on dance motion imitation have been realized [20], [12]. However, the robot dynamics have not been considered.

We propose a method for the imitation of whole-body motion for humanoid robots based on the stack of tasks framework. This method allows to quickly retarget a dynamic motion demonstrated by a human expert and to adapt the dynamics of the human body to the own dynamics of the robot. Then the output motion is modified or edited to rectify the differences with the original motion that were introduced by the previous retargeting. The obtained motion is dynamically consistent, and could be directly applied on the real humanoid robot. The motion generation method relies on an inverse-dynamics solver based on a cascade of quadratic programs [4]. Each quadratic program is associated to a desired task. The flexibility of the scheme allows the addition of arbitrary tasks on the joint space and operational space levels to rectify the movements. Fig. 2 illustrates the adopted method that we applied on a dancing motion. This dance motion is performed by a human and recorded with an optical motion capture system. This motion is then retargeted, edited and finally executed on the HRP-2 robot in simulation [5].

V. Conclusion

We presented our works relative to anthropomorphic motions. We performed task recognition, full-dynamic motion generation, motion retargeting and editing in a unified framework: the stack of tasks. Thanks to the genericity of the task function formalism, our works can be further extended. For example, for the recognition, the use of the task function formalism applied to human motion is currently investigated. Also, preliminary results on the real robot for the retargeting and editing method have been obtained.

References

[1] N. Mansard and F. Chaumette. Task sequencing for high level sensor-based control. *IEEE Transactions on Robotics*, 23(1):60–72, February 2007.

[2] C. Samson, M. Le Borgne, and B. Espiau. *Robot Control: the Task Function Approach*. Clarendon Press, Oxford, UK, 1991.

[3] S. Hak, N. Mansard, and O. Stasse. Humanoid robot task recognition from movement analysis. In *International Conference on Humanoid Robots (Humanoids)*, Nashville, USA, December 2010.

[4] L. Saab, N. Mansard, F. Keith, J.Y. Fourquet, and P. Soueres. Generation of dynamic motion for anthropomorphic systems under prioritized equality and inequality constraints. In *International Conference on Robotics and Automation (ICRA)*, Shanghai, China, May 2011.

[5] O. Ramos, L. Saab, S. Hak, and N. Mansard. Dynamic motion capture and edition using a stack of tasks. In *International Conference on Humanoid Robots (Humanoids)*, Bled, Slovenia, October 2011.

[6] B. Espiau, F. Chaumette, and P. Rives. A new approach to visual servoing in robotics. *IEEE Transactions on Robotics and Automation*, 8(3):313–326, 1992.

[7] B. Siciliano and J. J. Slotine. A general framework for managing multiple tasks in highly redundant robotic systems. In *IEEE International Conference on Advanced Robotics (ICAR)*, Pisa, Italy, June 1991.

[8] V. Krüger, D. Kragic, A. Ude, and C. Geib. The meaning of action: A review on action recognition and mapping. *Advanced Robotics*, 21(13):1473–1501, 2007.

[9] S. Schaal, A. J. Ijspeert, and A. Billard. Computational approaches to motor learning by imitation. *Philosophical Transaction of the Royal Society of London, series B*, 358(1431):537–547, March 2003.

[10] E. Drumwright and M. Matarić. Generating and recognizing free-space movements in humanoid robots. In *IEEE/RAS Intelligent Robots and Systems (IROS)*, Las Vegas, USA, October 2003.

[11] O. C. Jenkins and M. Matarić. Performance-derived behavior vocabularies: Data-driven acquisition of skills from motion. *International Journal of Humanoid Robotics*, 1(2):237–288, June 2004.

[12] S. Nakaoka, A. Nakazawa, F. Kanehiro, K. Kaneko, M. Morisawa, H. Hirukawa, and K. Ikeuchi. Learning from observation paradigm: Leg task models for enabling a biped humanoid robot to imitate human dances. *International Journal of Robotics Research*, 26(8):829–844, August 2007.

[13] R. Chalodhorn, K. MacDorman, and M. Asada. Humanoid robot motion recognition and reproduction. *Advanced Robotics*, 23:349–366, 2009.

[14] M. Mühlig, M. Gienger, J. Steil, and C. Goerick. Automatic selection of task spaces for imitation learning. In *IEEE/RAS Intelligent Robots and Systems (IROS)*, St Louis, USA, October 2009.

[15] S. Schaal J. Peters. Reinforcement learning of motor skills with policy gradients. *Neural Networks*, 21(4):682–697, May 2008.

[16] K. Miura, M. Morisawa, and S. Nakaoka. Robot motion remix based on motion capture data towards human-like locomotion of humanoid robots. In *International Conference on Humanoid Robots (Humanoids)*, pages 596–603, Paris, France, December 2009.

[17] M. Gleicher. Retargeting motion to new characters. In *ACM Conference on Computer graphics and interactive techniques (SIGGRAPH)*, pages 33–42, New York, NY, USA, 1998.

[18] N.S. Pollard, J.K. Hodgins, M.J. Riley, and C.G. Atkeson. Adapting human motion for the control of a humanoid robot. In *International Conference on Robotics and Automation (ICRA)*, pages 1390–1397, Washington D.C., USA, May 2002.

[19] A. Safonova, N.S. Pollard, and J.K. Hodgins. Optimizing human motion for the control of a humanoid robot. In *Adaptive Motion of Animals and Machines (AMAM)*, Kyoto, Japan, March 2003.

[20] M. Riley, A. Ude, and C.G. Atkeson. Methods for motion generation and interaction with a humanoid robot: Case studies of dancing and catching. In *Workshop on Interactive Robotics and Entertainment (WIRE)*, pages 35–42, Pittsburgh, USA, April 2000.