Signal Processing 108 (2015) 136–146

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Whole-body humanoid robot imitation with pose similarity evaluation

Jie Lei [a], Mingli Song [a,*], Ze-Nian Li [b], Chun Chen [a]

[a] Zhejiang University, Hangzhou 310027, PR China
[b] Simon Fraser University, Burnaby, BC, Canada, V5A1S6

ARTICLE INFO

Article history:
Received 13 May 2014
Received in revised form
24 July 2014
Accepted 17 August 2014
Available online 26 August 2014

Keywords:
Humanoid robot
Pose transfer
Imitation
Similarity metric

ABSTRACT

Imitation is considered to be a kind of social learning that allows the transfer of information, actions, behaviors, etc. Whereas current robots are unable to perform as many tasks as human, it is a natural way for them to learn by imitations, just as human does. With the humanoid robots being more intelligent, the field of robot imitation has getting noticeable advance.

In this paper, we focus on the pose imitation between a human and a humanoid robot and learning a similarity metric between human pose and robot pose. In contrast to recent approaches that capture human data using expensive motion captures or only imitate the upper body movements, our framework adopts a Kinect instead and can deal with complex, whole body motions by keeping both single pose balance and pose sequence balance. Meanwhile, different from previous work that employs subjective evaluation, we propose a pose similarity metric based on the shared structure of the motion spaces of human and robot. The qualitative and quantitative experimental results demonstrate a satisfactory imitation performance and indicate that the proposed pose similarity metric is discriminative.

1. Introduction

With the development of robotics, robots are getting much smarter than they used to, especially for humanoid robots. However, they are not ready to perform many tasks as naturally as human beings. Imitation is considered as an effective solution to the problem. Specifically, imitation is an advanced behavior whereby an individual observes and replicates the behaviors of others. Robots have replaced humans in the assistance of performing repetitive and dangerous tasks in some fields, such as construction industry, medical surgery, toxic substances cleaning and space exploration, where they can take advantage of imitating human to some degree.

Imitation is about generating stable humanoid movements from the human motions, an overview and computational approaches to this problem can be found in [1]. Many of the imitation researches focus on the upper body and employ complex system setting. In [2], an analytical method was proposed to transfer the upper body motion from human to humanoid robot. Riley et al. [3] made use of some colored marks on a human upper body in order to be abstracted by a vision system based on external cameras and a head-mounted one of a humanoid. These marks were used to estimate the angular range of some joints with a kinematic model of the human to perform the imitation. Similar to [3], with the help of 34 markers placed on the human upper body and 2 markers attached on a conductor stick, Ott et al. [4] applied the data obtained by a motion capture system to allow a humanoid

* Corresponding author.
E-mail addresses: ljaylei@zju.edu.cn (J. Lei), brooksong@zju.edu.cn (M. Song), li@sfu.ca (Z.-N. Li), chenc@zju.edu.cn (C. Chen).

http://dx.doi.org/10.1016/j.sigpro.2014.08.030
0165-1684/

robot to mimic the human motion regarding to a Cartesian control approach. Aleotti et al. [5] adopted neural networks to learn a mapping between the positions of a human arm and an industrial robot arm. Based on Aleotti's work, Stanton et al. [6] extended it to a humanoid robot by training a feed forward neural network with particle swarm optimization for each degree of freedom (DOF). In the data collecting process, a robot was used to lead a human operator through series of paired synchronized movements captured by a motion capture, which was time-consuming and tedious. As they were mentioned, in order to ensure robot stability, the position of the robot's ankles did not employ neural networks. Since the neural networks could not always output ideal angles, the robot, as a rigid body, was apt to lose its balance. Meanwhile, a unified neural network training for the whole body was infeasible, considering convergence trouble. Whereas training with separate networks would cause correlation loss among the DOFs. Other imitation researches are mainly dedicated to humanoid gait or walking movements [7–9]. In conclusion, existing works have the following limitations:

- Imitation of the upper body or a single part is insufficient to meet the needs of humanoid robot [2–5].
- With requiring motion capture equipment, it is expensive and inconvenient for general use and unnatural for human–robot interaction [5,6,10].
- Lack of balance control and the whole body control [3,4,6].
- The imitation results are not qualitatively evaluated [6,10,11].

After performing the pose imitation, another important issue is "how can we evaluate the imitation similarity between a robot slave and the master". In [11], Zuher et al. gave a subjective evaluation by taking persons to mark the quality of an imitation with bad, poor, fair, good and excellent. Other existing research efforts are basically concentrated on the pose similarity of a single agent. The simplest metric is $L2$ distance, which does not sufficiently utilize the data dependency between DOFs. In [12], different weights were learned for DOFs, in correspondence with the fact that some DOFs had more influences on determining the similarity. Chen et al. [13] proposed a new rich pose feature set to effectively encode the pose similarity by utilizing features on geometric relations among body parts. Based on the pose feature set, a distance metric was learned in a semi-supervised manner. By matching the related DOFs of a robot and a human, we can apply these methods to evaluate the imitation similarity. However, robots and humans are different in DOF dimensions and physical constrains, i.e., they have different motion spaces. It is inappropriate to compare them directly.

The problem we are facing here can be regarded as a metric learning problem. Learning a good distance metric in feature space is crucial in real-world applications. Good distance metrics are important to many computer vision tasks, such as image classification [14–16], content-based image retrieval [17,18] and their applications [19,20]. Many useful algorithms and ideas were proposed in these papers to combine multiple feature sets, such as high-order distance-based multiview stochastic learning (HD-MSL [14]) and semi-supervised multiview distance metric learning (SSM-DML [16]). In our case, we believe that the human poses and the humanoid robot poses have much in common for their highly similar skeleton structures. Their differences depend on the number of DOFs and physical constrains. As a consequence, the shared motion space between the two agents can be a good metric space to study the pose similarity.

This paper proposes a novel humanoid robot imitation framework with pose similarity metric learning between human pose and robot pose, using a consumer camera (the Microsoft Kinect) and a humanoid robot (the Aldebaran Nao H25). The proposed framework summarized in Fig. 1 adopts dynamic balance control with realtime imitation performance. A shared representation of both robot pose and human pose is learned to evaluate the imitation similarity. Both qualitative and quantitative experimental results demonstrate a satisfactory imitation performance and indicate that the proposed pose similarity evaluation is discriminative.

Our main contributions are the following: (a) we propose a novel framework to perform pose imitation on the whole body motions rather than the upper body. (b) We actively keep single pose balance and introduce transient poses to achieve smooth pose sequence balance. (c) We demonstrate how shared structure can provide a quantitative evaluation to define the similarity between a human pose and a robot pose.

2. Humanoid robot imitation

2.1. Pose representation

The Kinect consists of a RGB camera, a depth sensor and provides 3D human skeleton tracking at 30 frames per second. Based on the position data obtained, we can calculate 20 DOF angles listed in Table 1, which are angles between pairs of related vectors. For example,

$$\theta^{H}_{HeadPitch} = \langle DV(Pos_{Spine}, Pos_{ShoulderCenter}), DV(Pos_{ShoulderCenter}, Pos_{Head}) \rangle \quad (1)$$

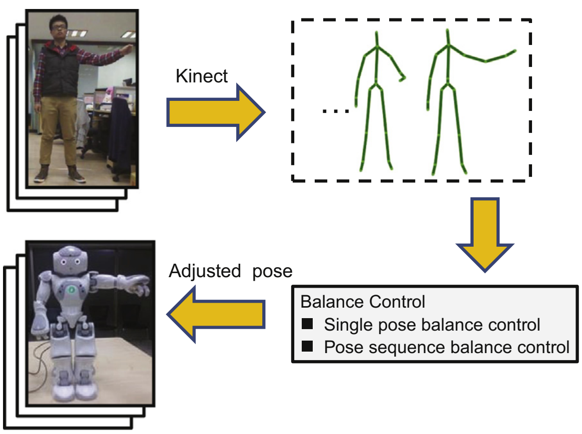

Fig. 1. Overview of our human to humanoid robot imitation system.