SCIENCE CHINA
Information Sciences

May 2019, Vol. 62 050205:1–050205:3
https://doi.org/10.1007/s11432-018-9667-0

• LETTER •
Special Focus on Human-Robot Hybrid Intelligence

Naturally teaching a humanoid Tri-Co robot in a real-time scenario using first person view

Liang GONG*, Xudong LI, Wenbin XU, Binhao CHEN, Zelin ZHAO, Yixiang HUANG & Chengliang LIU

School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, *China*

Received 31 August 2018/Accepted 19 October 2018/Published online 26 February 2019

Citation Gong L, Li X D, Xu W B, et al. Naturally teaching a humanoid Tri-Co robot in a real-time scenario using first person view. Sci China Inf Sci, 2019, 62(5): 050205, https://doi.org/10.1007/s11432-018-9667-0

Dear editor,

As a direct method of endowing robots with human knowledge, teaching renders the development of robot intelligence to an extraordinary extent. However, as we inevitably encounter complicated motions with multiple degrees of freedom (DOFs), traditional teaching methods continue to face a number of challenges.

Children learn through observing and reproducing adult behavior. In this natural way, they learn most effectively from the common sharing of comprehension of scenes and behavioral language of other humans. Hence, it falls on the shoulders of human-robot interaction (HRI) technology to create a highly effective teaching method. As a branch of HRI, natural teaching represents a kind of teaching paradigm that is user-friendly and coordinates human and robot in scene comprehension. This is especially true when it comes to employing the first person view (FPV), where the vision fields of the robot and the manipulator are synchronously aligned. Aimed at completing tasks with specific semantic information, natural teaching is a highly efficient end-to-end method for human-environment interaction. Training with such tasks is also conductive to establish a deep understanding of the potential implications from training data through subsequent intelligence algorithms, thus furthering the evolution of a high level of intellectual development [1, 2].

Here, we present a novel natural teaching paradigm that leverages the full potential to empower a humanoid Tri-Co (coexisting-cooperative-cognitive) robot from FPV, and facilitates manipulation intelligence and teleoperation [3, 4]. Steps begin with the establishment of a human-in-the-loop telepresence system to manipulate the robot from the FPV, engaging in a range of techniques in the humanoid robot setup, scene perception, motion capture and imitation. Human behavior then is recorded and imitated in real time in order to realize the robot's learning from demonstration. The result is examined through a delicate obstacle avoidance experiment in a cluttered background to validate its feasibility. To verify the natural teaching paradigm, an open-source 3D printing humanoid robot InMoov was employed [5]. With the ability to completely mimic human motion, the difference between human motion and robot imitation during motion synchronization can be easily accessed from the demonstration. For further explanation, 22 out of 29 DOFs are controlled during the motion teleoperation process, including 5 DOFs for each hand, 4 for each arm, 3 for each shoulder, and 2 for the neck [6].

As for scene perception, it is made possible to remotely perceive the complicated surroundings around the robot for the manipulator by visual feedback. With a camera installed in the eye of the robot, the manipulator can make decisions from the FPV via wearing virtual reality glasses. Intuitively, the FPV provides a more natural method

* Corresponding author (email: gongliang_mi@sjtu.edu.cn)

to teach robots because it avails itself to coordinate human-robot motions, align their vision, and in turn fuse the semantic understanding of the scene and corresponding behaviors.

For motion capture, a modular system composed of 32 9-axis wearable sensors is adopted. Human's real-time motion can be captured and reflected on a skeletal model using BVH (biovision hierarchy) data [7,8]. Subsequently, the BVH data is broadcasted through TCP so that the motion can be transmitted to the humanoid robot.

Mapping algorithm. In the imitation of human motion, the key point lies in sending corresponding joint angles computed from BVH data to the robot. BVH provides us with three euler angles for each node, from which it is possible to acquire the rotation matrix between child and parent links. Euler angles are denoted by a rotation order of ZYX as φ, θ, ψ, and the rotation matrix of a child frame with respect to a parent frame is given by

$$R_{\text{child}}^{\text{parent}} = \begin{pmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\psi & -\sin\psi \\ 0 & \sin\psi & \cos\psi \end{pmatrix}. \tag{1}$$

In this study, the primary concern for motion description lies in postures. Here, we consider human motion as a sequence of rotation matrices f_i:

$$f_i = \left\{ R_{\text{LHand}}^{\text{LForearm}}, R_{\text{LForearm}}^{\text{LArm}}, R_{\text{LArm}}^{\text{Body}}, R_{\text{Head}}^{\text{Body}}, R_{\text{RHand}}^{\text{RForearm}}, R_{\text{RForearm}}^{\text{RArm}}, R_{\text{RArm}}^{\text{Body}} \right\}. \tag{2}$$

As shown above, each posture is described by defining it as a sequence of rotation matrices at time i, i.e., $R_{\text{LHand}}^{\text{LForearm}}$ stands for the rotation matrix between left hand and left forearm. Likewise, robot motion is defined as another sequence. Our goal in using such a process is to eliminate the difference between each corresponding rotation matrix of human and robot to the greatest extent. Due to biological constraints, human bodies cannot have three rotational DOFs at each joint, and not all of them are independent. Moreover, with mechanical constraints, some joints of humanoid robots are also unable to rotate in three independent directions. Considering such comparability, each joint is assigned with a specific mapping algorithm. Thanks to structural symmetry, the algorithms for $R_{\text{LJoint1}}^{\text{LJoint2}}$ and $R_{\text{RJoint1}}^{\text{RJoint2}}$ share the same principle.

Mapping between shoulders. The first case entails the conversion from three human DOFs to three robot DOFs. Three rotational joints were installed on each part of the shoulder of InMoov. The axes of rotation can be approximately treated as perpendicular to each other. Let α, β, γ denote the joint angles of the three shoulder joints, and the rotation matrix of the arm link with respect to the body can be similarly expressed as

$$R_{\text{Arm}}^{\text{Body}} = \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\gamma & -\sin\gamma \\ 0 & \sin\gamma & \cos\gamma \end{pmatrix}. \tag{3}$$

With (1) and (3), we can derive a one-to-one correlation between the (φ, θ, ψ) and (α, β, γ).

Mapping between elbow joints. The second case entails the conversion from two human DOFs to one robot DOF. Compared with human elbows that can bend and rotate, those of the robot are not able to rotate. Hence, we only need to compute the joint angle for bending. Define the angle as Ω. With the assumption that sensors are fixed with respect to the human body, and the x-direction is along the forearm link, we can derive the following equations:

$$\hat{\boldsymbol{x}}_2^2 = (1, 0, 0)^{\mathrm{T}}, \tag{4}$$

$$\hat{\boldsymbol{x}}_2^1 = R_2^1 \hat{\boldsymbol{x}}_2^2 = (\cos\varphi\cos\theta, \cos\varphi\sin\theta, -\sin\theta)^{\mathrm{T}}, \tag{5}$$

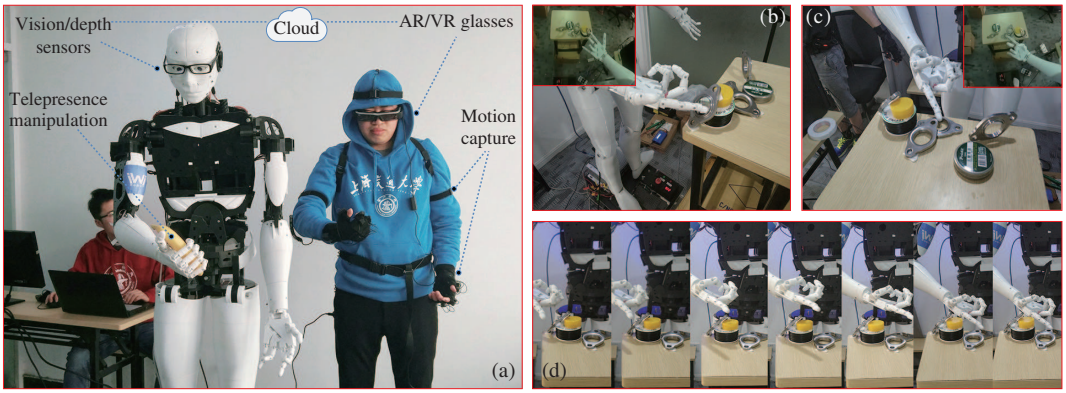
$$\Omega = \pi - \langle \hat{\boldsymbol{x}}_2^1, \hat{\boldsymbol{x}}_1^1 \rangle = \pi - \arccos(\cos\varphi\cos\theta). \tag{6}$$

R_2^1 stands for the rotation matrix of frame $x_2y_2z_2$ with respect to $x_1y_1z_1$. $\hat{\boldsymbol{x}}_1^1$ is a unit vector of $\boldsymbol{x}_1$ in frame $x_1y_1z_1$.

Mapping between neck joints. The third case entails the conversion from three human DOF (φ, θ, ψ) to two robot DOF (α, β). With mechanical constraints of robots, rotation in one direction has to be abandoned. In this way, α and β of the robot can be resembled by φ and θ.

Natural teaching. The process of natural teaching is shown in Figure 1(a). First, a vision sensor is employed to project the mission scene onto the VR glasses. The motion of a human is then captured by motion perception with a set of wearable sensors, which then presents the collected motion data in a BVH format. Later, motion data is transmitted to an industrial PC (IPC) connected to the robot through a TCP/IP or cloud server and parsed according to the BVH format, which is followed by converting the parsed Euler angles to corresponding joint angles through a fast mapping algorithm and encapsulating it in a communication protocol. Finally, the IPC sends joint angles to the slave controller to control the robot.

In a testament to the feasibility of natural teaching from FPV, a delicate obstacle avoidance experiment was designed where the robot was remotely



Frame number	1	2	3	4	5	6	7	Average (cm)
Error (cm)	4.455	3.059	1.395	4.508	5.045	3.703	1.878	3.435

(e)

Figure 1 Natural teaching process. (a) Humanoid natural teaching system; (b) task starting point; (c) task end point; (d) frames during teaching process; and (e) error during teaching process.

operated to perform fast collision-avoidance motion. First, a cluttered obstacle scene was constructed. A demonstrator was required to bypass complex obstacles as close as possible. The robot's index finger was to be moved from the initial state (as shown in Figure 1(b)) to the final state (Figure 1(c)). Additionally, the experiment was required to be performed fast and coherently without collision and retreating. Such natural teaching experiments demonstrate that the operator can drive the robot remotely to perform complex tasks both efficiently and quickly. Furthermore, to realize the robot's learning from demonstration, the task solution was recorded in real-time.

The teaching control error is defined as the nearest distance of the robot end to the obstacle surface during task execution. The teaching control error was generated by the robot system stability deviation, the operator's unconscious jitter, and the amount of redundant drive provided for fast obstacle avoidance. The error accurately describes the operability of the natural teaching under fast teaching conditions, as the task is executed fast, coherently, and in a single period of time. The errors in a teaching process are shown in Figure 1(e). The average error was 3.435 cm, which could be reduced at a slow pace but could not be neglected for precise motion control. However, the error is acceptable in a life-size robot action scenario.

Conclusion. In this study, we presented a novel natural teaching paradigm for humanoid robots using FPV. To verify the effectiveness of a natural teaching paradigm, we constructed a human-in-the-loop telepresence system as the platform. The outcome of the delicate obstacle avoidance experiment demonstrated that natural teaching is particularly effective in imitating large-scale movement and complex motions with inferior precision. By the most natural means, the FPV-based teaching approach paves a new way for training a robot to cope with a dynamic environment through demonstration and autonomous learning.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant No. 51775333).

References

1 Wachter M, Asfour T. Hierarchical segmentation of manipulation actions based on object relations and motion characteristics. In: Proceedings of International Conference on Advanced Robotics, 2015. 549–556
2 Lim G H. Two-step learning about normal and exceptional human behaviors incorporating patterns and knowledge. In: Proceedings of IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, 2017. 162–167
3 Ding H, Yang X, Zheng N, et al. Tri-Co Robot: a Chinese robotic research initiative for enhanced robot interaction capabilities. Natl Sci Rev, 2018, 5: 799–801
4 Argall B D, Chernova S, Veloso M, et al. A survey of robot learning from demonstration. Robot Autonom Syst, 2009, 57: 469–483
5 Langevin G. Inmoov. 2014. http://www.inmoov.fr/project
6 Gong L, Gong C, Ma Z, et al. Real-time human-in-the-loop remote control for a lifesize traffic police robot with multiple augmented reality aided display terminals. In: Proceedings of the 2nd International Conference on Advanced Robotics and Mechatronics (ICARM), 2017. 420–425
7 Meng X, Pan J, Qin H. Motion capture and retargeting of fish by monocular camera. In: Proceedings of International Conference on Cyberworlds, 2017. 80–87
8 Dai H, Cai B, Song J, et al. Skeletal animation based on bvh motion data. In: Proceedings of the 2nd International Conference on Information Engineering and Computer Science, 2010. 1–4