SCIENCE CHINA
Information Sciences

• LETTER •

Naturally teaching a Humanoid Tri-Co Robot in a Real-time Scenario from First Person View

Liang GONG[1*], Xudong LI[1], Wenbin XU[1], Binhao CHEN[1]
Zelin ZHAO[1], Yixiang HUANG[1] & Chengliang LIU[1]

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

Citation Gong L, Li X D, Xu W B, et al. Naturally teaching a Humanoid Tri-Co Robot in a Real-time Scenario from First Person View. Sci China Inf Sci, for review

Dear editor,
As a direct way to endow industrial robots with humans knowledge, teaching renders the intelligence development of robots more than possible. However, traditional teaching methods are faced with difficulties when teaching is performed for complicated motions with multiple DOF (degree of freedom). And it is hard to realize knowledge accumulation of robots through traditional teaching in the absence of behavior recognition and semantic classification [1, 2].

Considering childrenś learning process, they observe the behavior of adults, and then reproduce it. Such a process is always natural and highly effective because human beings share the comprehension of scenes and behavioral language. Indeed, natural teaching is a branch of human-robot interaction (HRI) technology, representing a kind of teaching paradigm which is user-friendly and coordinates human and robot in scene comprehension. Aimed at completing tasks with specified human semantic information, natural teaching is an end-to-end and highly efficient method for interaction with the surroundings. Moreover, training with such tasks is conducive to establish a deep understanding of potential implications from training data through subsequent intelligence algorithms, thus achieving a high level of intellectual development.

In this letter, we present a novel natural teaching paradigm which leverages the full potential to empower a tri-co (Coexisting-Cooperative-Cognitive) humanoid robot from first person view(FPV), and facilitates the manipulation intelligence and teleoperation [3]. First, a human-in-the-loop telepresence system is built for the purpose of manipulating the robot from the FPV, with details in the humanoid robot setup, the scene perception, the motion capture and imitation. Next, human behavior is recorded and imitated in real time in order to realize the robots learning from demonstration [4]. Finally, a delicate obstacle avoidance experiment in a cluttered background is conducted to validate the feasibility of the proposed method.

At first, we employ humanoid robots as a platform to verify the natural teaching paradigm. The robot is established based on InMoov [5], an open-sourced 3D printing humanoid robot which can serve as a direct and natural platform for natural teaching. As they can completely reflect human motion, demonstrators can easily assess the difference between human motion and robot imitation during motion synchronization. During motion teleoperation process, 22 out of 29 DOF are controlled, including 5 DOF for each hand, 4 for each arm, 3 for each shoulder and 2 for the neck [6].

In the aspect of scene perception, the visual feedback makes it possible for the manipulator to perceive the complicated surroundings around the

[*]This invited submission is partially published in the conference IEEE ICARM2018 with the title of Human-robot Interaction Oriented Human-in-the-loop Real-time Motion Imitation on a Humanoid Tri-Co Robot.
[*] Corresponding author (email: gongliang_mi@sjtu.edu.cn)

robot remotely. Since the camera is installed in one eye of the robot, the manipulator wearing VR glasses can make decisions about movements from a first-person view. To build such real-time telepresence system, Raspberry Pi is selected as the processing unit to drive the Pi camera for remote live video monitoring. Then FFmpeg (Fast Forward mpeg) is adopted to convert the H.264 video which is obtained from raspberry to the MPEG1 format. Hence, the video stream can be received and decoded by VR glasses with better compatibility and real-time performance.

To capture manipulators motion information, a modular system composed of 32 9-axis wearable sensors is adopted. The humanś real-time motion can be captured and reflected on the skeletal model through BVH data [7,8]. Whatś more, the BVH data can be broadcasted through TCP, so that the motion can be transmitted to the humanoid robot.

Mapping Algorithm. To make the robot imitate human motion, the key point is to send corresponding joint angles computed from BVH data. BVH has provided us with three euler angles for each node, enabling us to ascertain the rotation matrix between child and parent links. Denote euler angles with a rotation order of ZYX as φ, θ, ψ, the rotation matrix of child frame with respect to parent frame is

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

In this work, postures are mainly concerned for motion description. Hence, we consider human motion as a sequence of rotation matrices f_i

$$\begin{aligned} f_i = \Big\{ & R_{LHand}^{LForearm}, R_{LForearm}^{LArm}, R_{LArm}^{Body}, R_{Head}^{Body}, \\ & R_{RHand}^{RForearm}, R_{RForearm}^{RArm}, R_{RArm}^{Body} \Big\} \end{aligned} \tag{2}$$

Thus, each posture is defined as a sequence of rotation matrices at time i, i.e., $R_{LHand}^{LForearm}$ stands for the rotation matrix between left hand and left forearm. Similarly, we can also define robot motion as another sequence. The goal is to eliminate the difference between each corresponding rotation matrix of human and robot as much as possible. Human, with biological constraints, cannot have 3 rotational DOF at each joint and some of them are not completely independent. Due to mechanical constraints, some joints of humanoid robot are unable to rotate in three independent directions either. Hence, each joint is assigned a specific the mapping algorithm. Thanks to the structural symmetry, the algorithms for $R_{LJoint1}^{LJoint2}$ and $R_{RJoint1}^{RJoint2}$ share the same principle.

Mapping between shoulders: The first case entails conversion from 3 human DOF to 3 robot DOF. Three rotational joints are installed on each shoulder part of InMoov and their axes of rotation can be approximately treated as perpendicular to each other. Denote the joint angles of 3 shoulder joints as respectively α, β, γ and the rotation matrix of the arm link with respect to the body can be similarly expressed as

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

With Equ.1 and Equ.3, we can derive a one-to-one correlation

$$\alpha = \ \varphi, \ \beta = \ \theta, \ \gamma = \ \psi \tag{4}$$

Mapping between elbow joints:The second case entails the conversion from 2 human DOF to 1 robot DOF. Human elbows are able to bend and rotate while those of the robot can only bend. Then we need to compute the joint angle for bending, which is Ω.With the assumptions that sensors are fixed with respect to human body and the x-direction is along the forearm link, we can derive the following equations.

$$\hat{\boldsymbol{x_2}}^2 = (1,0,0)^T \tag{5}$$

$$\hat{\boldsymbol{x_2}}^1 = R_2^1 \hat{\boldsymbol{x_2}}^2 = (cos\varphi cos\theta, cos\varphi sin\theta, -sin\theta)^T \tag{6}$$

$$\Omega = \pi - < \hat{\boldsymbol{x_2}}^1, \hat{\boldsymbol{x_1}}^1 > = \pi - arccos(cos\varphi cos\theta) \tag{7}$$

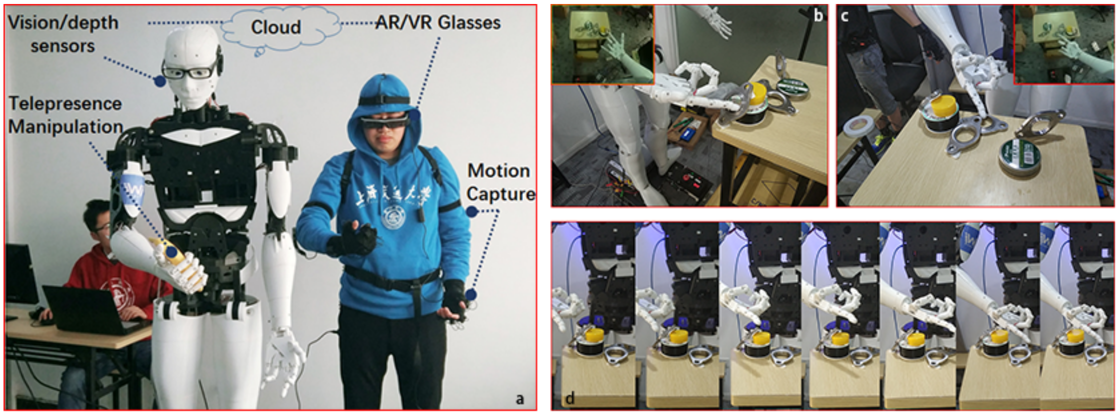
R_2^1 stands for the rotation matrix of frame $x_2 y_2 z_2$ with respect to $x_1 y_1 z_1$. $\hat{\boldsymbol{x_1}}^1$ is a unit vector of $\boldsymbol{x_1}$ in frame $x_1 y_1 z_1$.

Mapping between neck joints:The third case requires the conversion from 3 human DOF(φ, θ, ψ) to 2 robot DOF (α, β).Due to the mechanical constraints, rotation in one direction has to be abandoned. The solution to this case resembles that for the shoulder joint and be written as

$$\alpha = \ \varphi, \beta = \ \theta \tag{8}$$

Natural teaching. The process of natural teaching is shown in Fig. 1(a). First, a vision sensor is employed to project the mission scene onto the VR glasses. Second, motion perception captures the motion of human with a set of wearable sensors and presents the collected motion data in BVH (BioVision Hierarchy) format. Then motion data are transmitted to an industrial PC (IPC) which is installed on the robot side through TCP/IP or Cloud Server and parsed according to BVH format. Next, the parsed euler angles are converted to corresponding joint angles through a fast mapping algorithm and encapsulated in a communication protocol. At last, IPC sends joint angles to the slave controller to control the robot.

In order to verify the feasibility of the natural teaching from first person view, we designed a delicate obstacle avoidance experiment in which the



Frame No.	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 (e) Error during teaching process.

demonstrator remotely operated the robot to perform fast collision avoidance motion. First we constructed a cluttered obstacle scene and request the demonstrator to drive the robot to bypass complex obstacles as close as possible.The demonstrator must move the robot's index finger from the initial state (as shown in Fig. 1(b)) to the final state (Fig. 1(c)).Experiment requires the demonstrator to perform this task fast and coherently without collision and retreating. The natural teaching experimental process (as shown in Fig. 1(d)) shows that demonstrator can drive robot to perform complex task efficiently and quickly. Moreover, the task solution data is recorded in real time in order to realize the robots learning from demonstration.

We define the nearest distance of the robot end to the obstacle surface during task execution as the teaching control error. This error comes from the robot system stability deviation, the operator's unconscious jitter, and the amount of redundant drive provided for fast obstacle avoidance. Since the task is executed in one-time, fast and coherently, the error can well characterize the operability of the natural teaching under fast teaching conditions. The errors in a teaching process is shown in Fig. 1(e). The average error is 3.435cm which cannot be ignored for precise motion control although it can be reduced at a slow pace.However,the error is acceptable in a life-size robot action scenario.

Conclusion. In this letter, we present a novel natural teaching paradigm for humanoid robot from FPV. A human-in-the-loop telepresence system is built as a platform to verify the effectiveness of natural teaching. The result of delicate obstacle avoidance experiment shows 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 dynamic environment through demonstration and autonomous learning.

Acknowledgements This study was supported by the Natural Scientific Foundation of China under Grant NO.51775333.

References

1 M. Wachter and T. Asfour, Hierarchical segmentation of manipulation actions based on object relations and motion characteristics, in International Conference on Advanced Robotics, 2015, pp. 549-556.
2 G. H. Lim, Two-step learning about normal and exceptional human behaviors incorporating patterns and knowledge, in IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, 2017, pp. 162-167.
3 H. Ding, X. Yang, N. Zheng, M. Li, Y. Lai, and H. Wu, Tri-co robot: a chinese robotic research initiative for enhanced robot interaction capabilities, National Science Review, p.nwx148, 2017. [Online]. Available: http://dx.doi.org/10.1093/nsr/nwx148
4 B. D. Argall, S. Chernova, M. Velso, and B. Browning, A survey of robot learning from demonstration, Robotics & Autonomous Systems,vol.57, no.5, pp.469-483, 2009.
5 G. Langevin,Inmoov, http://www. inmoov.fr/project, 2014
6 L. Gong, C. Gong, Z. Ma, L. Zhao, Z. Wang, X. Li, X. Jing, H. Yang,and C. Liu, Real-time human-in-the-loop remote control for a lifesize traffic police robot with multiple augmented reality aided display terminals, in 2017 2nd International Conference on Advanced Robotics and Mechatronics (ICARM), Aug 2017, pp.420-425.
7 X. Meng, J. Pan, and H. Qin, Motion capture and retargeting of fish by monocular camera, in International Conference on Cyberworlds, 2017, pp.80-87.
8 H. Dai, B. Cai, J. Song, and D. Zhang, Skeletal animation based on bvh motion data, in 2010 2nd International Conference on Information Engineering and Computer Science, Dec 2010, pp. 1C4.