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 human's knowledge, teaching renders the intelligence development of robots to an extraordinary extent. However, as we inevitably encounter complicated motions with multiple DOFs (degree of freedom), traditional teaching methods are in face of difficulties. Moreover, the object of knowing knowledge accumulation of robots becomes also impracticable by traditional teaching since behavior recognition and semantic classification is absent.

Children learn through observing adult behavior and reproducing it. In this natural way they learn most effectively due to the common sharing of comprehension of scenes and behavioral language of human beings. Hence, it falls on the shoulder of human-robot interaction (HRI) technology a task as to how we achieve a teaching method of high effectiveness. As a branch of HRI, natural teaching represents a kind of teaching paradigm which is user-friendly and coordinates human and robot in scene comprehension. Aimed at completing tasks with specific human semantic information, natural teaching is a highly efficient end-to-end method for human-environment interaction. Training with such tasks is also conductive in establishing a deep understanding of potential implications from training data through subsequent intelligence algorithms, thus further achieving a high level of intellectual development.

Here, 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. Steps begin with the establishment of a human-in-the-loop telepresence system to manipulate the robot from the FPV, engaging a range of techniques in humanoid robot setup, scene perception, motion capturing and imitation. Then, human behavior 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 in validation of feasibility.

To verify the natural teaching paradigm, we first employed humanoid robots as the platform. The robot is established based on InMoov [5], an open-sourced 3D printing humanoid robot served as a direct and natural platform for natural teaching. With the ability to completely reflect human motion, the difference between human motion and robot imitation during motion synchronization can be easily accessed in demonstration. For further explanation, 22 out of 29 DOF are controlled during motion teleoperation process, including 5 DOF 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

*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)

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 FPV wearing VR. To realize such real-time telepresence system, we selected Raspberry Pi for processing to drive he Pi camera for remote live video monitoring, accompanied by FFmpeg (Fast Forward mpeg) to convert the H.264 video obtained by Raspberry Pi to the MPEGI format. In this way, we made the video stream receivable and decodable by VR glasses with real-time performance in better compatibility.

For motion capturing, a modular system composed of 32 9-axis wearable sensors is adopted. Humanś real-time motion can be captured and reflected on the skeletal model through BVH data [7, 8]. Then, BVH data is broadcasted through TCP in order that the motion can be transmitted to the humanoid robot.

Mapping Algorithm. In imitation of human motion, our 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 we acquire 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

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

Main concern for motion description lies in postures in this work. Here we consider human motion as a sequence of rotation matrices f_i

$$f_i = \left\{ R_{LHand}^{LForearm}, R_{LForearm}^{LArm}, R_{LArm}^{Body}, R_{Head}^{Body}, R_{RHand}^{RForearm}, R_{RForearm}^{RArm}, R_{RArm}^{Body} \right\} \quad (2)$$

As shown above, we describe each posture by defining it 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. Likewise, we define robot motion as another sequence. Our goal of such process is eliminating the difference between each corresponding rotation matrix of human and robot to the most extent. Due to biological constraints, human bodies cannot have 3 rotational DOF at each joint and not all of them are independent. Also, with mechanical constraints, some joints of humanoid robot are unable to rotate in three independent directions, either. Considering such comparability, we assigned each joint with 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. The 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

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

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

$$\alpha = \varphi, \ \beta = \theta, \ \gamma = \psi \quad (4)$$

Mapping between elbow joints: The second case entails the conversion from 2 human DOF to 1 robot DOF. Compared with human elbows that are able to bend and rotate, those of the robot lack the ability of rotation. Hence, we need only compute the joint angle for bending. Define the angle as 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.

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

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

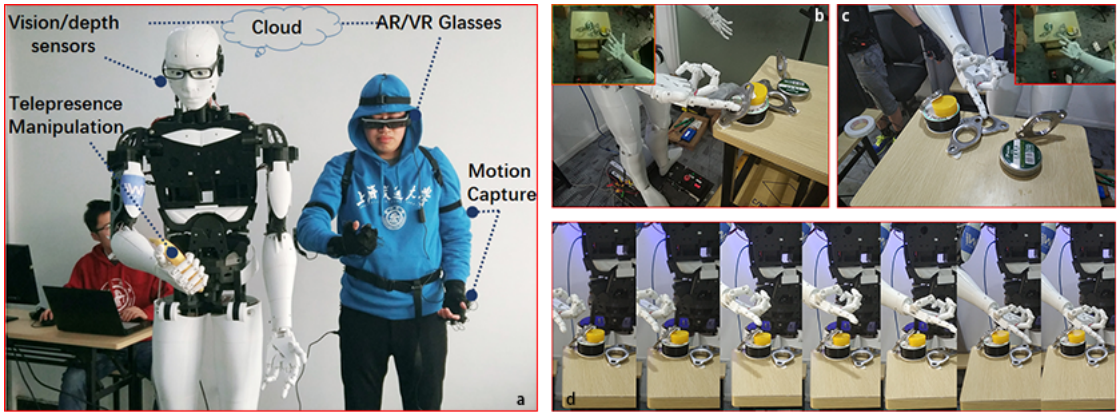
$$\Omega = \pi - < \hat{\boldsymbol{x_2}}^1, \hat{\boldsymbol{x_1}}^1 >= \pi - arccos(cos\varphi cos\theta) \quad (7)$$

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 requires the conversion from 3 human DOF(φ, θ, ψ) to 2 robot DOF (α, β). With mechanical constraints of robots, rotation in one direction has to be abandoned. In this way, the solution to this case which resembles the shoulder joint can be written as

$$\alpha = \varphi, \beta = \theta \quad (8)$$

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. Then, motion of human ia captured by motion perception with a set of wearable sensors, which then presents the collected motion data in BVH (BioVision Hierarchy) format. Later, motion data is transmitted to an industrial PC (IPC) installed on the robot side through TCP/IP or Cloud Server and parsed according to BVH format, followed by converting the parsed euler angles to corresponding joint angles through a fast



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.

mapping algorithm and encapsulated in a communication protocol. Finally, IPC sends joint angles to the slave controller to control the robot.

In testament of the feasibility of natural teaching from FPV, a delicate obstacle avoidance experiment is designed where the robot is remotely operated to perform fast collision avoidance motion. To begin with, a cluttered obstacle scene is constructed. Demonstrator is required to bypass complex obstacles as closely as possible. The robot's index finger is to be moved from the initial state (as shown in Fig. 1(b)) to the final state (Fig. 1(c)). Additionally, the experiment is required to be performed fast and coherently without collision and retreating. Such process of natural teaching experiment demonstrates that operator can drive the robot remotely to perform complex tasks both efficiently and quickly. Also, as to realize the robot's learning from demonstration, the task solution is 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 is 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 well describes the operability of the natural teaching under fast teaching conditions, for the task is executed fast, coherently and in one-time. The errors in a teaching process is shown in Fig. 1(e). The average error is 3.435cm, which is able to be reduced at a slow pace but cannot be neglected for precise motion control. 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. To verify the effectiveness of natural teaching paradigm, we constructed a human-in-the-loop telepresence system as the platform. Outcome of the delicate obstacle avoidance experiment has 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 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.