Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Fast and Intuitive Kinematics Mapping for Human-Robot Motion Imitating: A Virtual-Joint-Based Approach ⋆

Ziwei Wang * **Rongjian Liang** ** **Zhang Chen** *,*** **Bin Liang** *,***

* *Department of Automation, Tsinghua University, Beijing, 100084, China (e-mail: wang-zw16@mails.tsinghua.edu.cn, {cz_da, bliang, taozhang}@tsinghua.edu.cn).*
** *Department of Computer Scicence and Engineering, Texas A&M University, College Station, Texas, 77840, USA (e-mail: liangrj14@tamu.edu).*
*** *Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China.*

Abstract: It is quite difficult to imitate the motion of human arms using non-humanoid robots due to their dissimilar embodiments (degree-of-freedom, body morphology, and constraints). However, in most cases of the robotic imitation, the human operator and the robot would not share the same kinematic configuration. This paper addresses the motion imitation problem between the human arm and an industrial robot, where a commonly-used UR5 robot is considered. The motion of the human arm is obtained by an inertial motion capture system, and then the captured motion is reproduced using the UR5 embodiment. A virtual-joint-based approach is proposed to facilitate the fast and intuitive kinematics mapping between the human arm and the UR5 robot, leading to a robotic imitation system that can imitate the tip location and configuration of the human arm simultaneously. The proposed approach is verified experimentally on a real UR5 robot and compared with classic Cartesian-space-based mapping approach and joint-space-based approach.

Keywords: Robotic imitation, kinematics mapping, dissimilar embodiment, programming by demonstration, motion similarity.

1. INTRODUCTION

Robotic imitation is an effective tool in robotic programming and teleoperation Wang et al. (2020); Wang et al. (2019a,b). Robots can learn from the demonstration of human operators or other robots, which is often referred to learning by imitation or programming by demonstration Alissandrakis et al. (2007), Argall et al. (2009), to avoid intricate code programming. Also, programming by demonstration is more intuitive such that it can promote natural and efficient knowledge transfer from human to robots or between robots.

One of key challenges in robotic imitation is kinematics mapping Xiaojun Zhao et al. (2004), especially across dissimilar bodies. Kinematics mapping is the process of transforming the captured motion of the demonstrator into the imitators own capacity Nehaniv and Dautenhahn (2000). Many existing robotic imitation systems employed humanoid robots Calinon and Billard (2007), Lin et al. (2014)), which have similar bodies with their demonstrators, such as the human arm, so that kinematics mapping can be directly implemented by mapping the joint states from the demonstrator to the imitator. However, most of the robots may have different embodiments from humans', including industrial robots, space robots and some humanoid robots. Dissimilar embodiments result in great challenges in terms of kinematics mapping. Li et al. (2017) fused the human motion information captured by the Kinect and inertial measurement device, and then reproduced the motion using the humanoid Baxter Research Robot. Kormushev et al. (2011) used a free-standing humanoid robot to imitate the whole bodily motion of human beings. A simplified geometric approach Shahverdi and Masouleh (2016) was developed to solve the inverse kinematic problem of the upper-body of NAO robots. A whole-body imitation system was developed in Zhang et al. (2016) by fitting corresponding eight kinematic chains between the NAO robot and human body. Suleiman et al. (2008) proposed an optimization framework to generate the upper body motion of humanoid robots from human captured motions, taking the humanoid physical capabilities into account. A robotic imitation system using a small-size humanoid robot–Darwin-OP was developed in Lee et al. (2012). Hwang et al. (2014) developed a humanoid

⋆ This research was supported by the National Natural Science Foundation of China (Grant No. 61673239, 61703228); Science and technology project of Shenzhen (Grant No. JCYJ20160428182227081, JCYJ20160301100921349); Science and Technology Planning Project of Guangdong Province (Grant No. 2017B010116001). (Corresponding author: Zhang Chen)

robot imitation system, in which the neural-network-based inverse kinematics (IK) was utilized to reduce the computation time and improve the tracking accuracy. Besides the humanoid robots, there are also researches on the imitation of industrial robots, for example, the demonstration in robotic assembly Zhu and Hu (2018). Laguillaumie et al. (2016) proposed an integrated approach to the problem of reproducing a human hand or tool motion task with a robot. Jha et al. (2017) proposed a simplified approach of imitation learning for an industrial robot, realized an end-effector trajectory planning, with considering the features of the human arm kinematic model.

The main contribution of this paper is the development of an on-line imitation system for the human arm and an industrial robot (UR5), aiming at location of the human wrist and imitating the configuration of the human arm. The motion of human arm is obtained with an inertial motion capture system, and then the captured motion is reproduced by the UR5 robot. A virtual-joint-based method is applied to realize fast and intuitive kinematics mapping between the human arm and UR5. Experiments are conducted on a UR5 robot and the performance of the proposed kinematics mapping approach is compared with the classic Cartesian-space-based mapping approach and joint-space-based approach.

2. INERTIAL MOTION CAPTURE SYSTEM

The modular perception neuron system is used to capture the human motion in our research. It generates the Biovision Hierarchy (BVH) data, which records the human pose as a nested-structure of parent-joints and child-joints, containing information that specifies the location of the skeletal joint relative to its parent-joint. The upper-limb of human beings mainly consists of the shoulder complex, the elbow complex and the wrist joint Gopura and Kiguchi (2009), whose motions are shown in Fig. 1. Basically, the shoulder joint allows 3 DOFs motions, i.e., shoulder abduction/adduction, shoulder flexion/extension and internal/external rotation. The elbow complex allows 2 DOFs motions, i.e., elbow flexion/extension and supination/pronation. The wrist joint allows 2 DOFs motions, i.e., wrist flexion/extension and radial/ulnar deviation. In total, the human upper-limb can be modeled as a 7-DOF structure.

To facilitate further use of human data, the Shoulder-frame, Elbow-frame and Wrist-frame are attached to the human arm in Fig. 2. The shoulder frame moves with the shoulder motion, the elbow frame with the elbow flexion/extension and the wrist frame with the elbow flexion/extension, wrist flexion/extension and radial/ulnar deviation. In this way, the motion of the human arm can be represented by the displacement and orientation of these frames, which can be obtained via the relative location specifications in BVH data. l_1 and l_2 are the length of the human upper-arm and forearm, respectively. The states of the human arm can be represented as $S^D = \begin{bmatrix}\theta_1^D & \theta_2^D & \theta_3^D & \theta_4^D & \theta_5^D & \theta_6^D & \theta_7^D\end{bmatrix}^T$.

3. KINEMATICS OF THE UR5 ROBOT

An UR5 robot is used as the imitator of the human arm with 6 links Kufieta (2014). In order to obtain the relation between cascade links, the Denavit-Hartenberg (DH) convention is adopted, where the states of UR5 are expressed as $S^I = \begin{bmatrix}\theta_1^I & \theta_2^I & \theta_3^I & \theta_4^I & \theta_5^I & \theta_6^I\end{bmatrix}^T$. The UR5 robot is mounted on a metal holder and the orientation of the base link-frame of the UR5 is shown in Fig. 3. The roll, pitch and yaw (RPY) Euler-angle from the world frame to the base link-frame of the UR5 is $(\pi/2, 0, \theta_m)$, where θ_m is an optimization variable to enhance the imitation performance, which will be explained in the next section. The initial configuration is set as ${}^0S^I = [0, \pi, 0, 0, \frac{\pi}{2}, 0]^T$.

4. VIRTUAL-JOINT-BASED KINEMATICS MAPPING

4.1 General Idea

It is noted that the embodiments of human arm and the UR5 robot are dissimilar, leading to challenges in the kinematics mapping process. To be specific, the human arm has 7 DOFs while a UR5 robot has 6 DOFs. Furthermore, with different body morphology, it is difficult to find a satisfactory kinematics mapping solution even though one of the DOFs of human arm is ignored.

A virtual-joint-based approach is proposed for the first time in our previous work Chen et al. (2020) to tackle the kinematics mapping across dissimilar embodiments. Firstly, the DOFs of the demonstrator (human arm) and the imitator (UR5 robot) are divided into the same number of groups. A set of DOFs is called one virtual joint, and a virtual joint of the demonstrator and the corresponding one of the imitator form a virtual joint pair. Then metrics of motion similarity can be defined at the virtual-joint-level. It means that diverse types of metrics can be employed to measure the similarities between the virtual joints in one pair. And the kinematics mapping solution can be found by minimizing the metrics for all virtual joint pairs. the method is proposed as a general frame of kinematics mapping between robots with similar/dissimilar configurations. It should be noted that the key of the method is the virtual joint. With different design of the virtual joint, the method can be applied in different cases.

In the developed system, the DOFs of the human arm and the UR5 are divided into three groups respectively.

$$S^D = \begin{bmatrix}S^{D_1 T} & S^{D_2 T} & S^{D_3 T}\end{bmatrix}^T \tag{1}$$

where $S^{D_1} = \begin{bmatrix}\theta_1^D & \theta_2^D & \theta_3^D\end{bmatrix}^T$, $S^{D_2} = \begin{bmatrix}\theta_4^D\end{bmatrix}^T$, and $S^{D_3} = \begin{bmatrix}\theta_5^D & \theta_6^D & \theta_7^D\end{bmatrix}^T$.

$$S^I = \begin{bmatrix}S^{I_1 T} & S^{I_2 T} & S^{I_3 T}\end{bmatrix}^T \tag{2}$$

where $S^{I_1} = \begin{bmatrix}\theta_1^I & \theta_2^I\end{bmatrix}^T$, $S^{I_2} = \begin{bmatrix}\theta_3^I\end{bmatrix}^T$, and $S^{I_3} = \begin{bmatrix}\theta_4^I & \theta_5^I & \theta_6^I\end{bmatrix}^T$.

Consequently, three virtual joint pairs are formed. Then metrics of motion similarity can be defined. As pictured in Fig. 4, free vector v_{link2} is considered from the origin of the frame1 to that of the frame2 of the UR5, v_{link3} from the frame2 to the frame3, $v_{upper-arm}$ from the Shoulder-frame to the Elbow-frame, and $v_{forearm}$ from the Elbow-frame frame to the Wrist-frame. θ_e is the angle down from v_{link2} to v_{link3}, and θ_4^D is the angle down from $v_{upper-arm}$ to $v_{forearm}$. The desired global performance is that v_{link2}

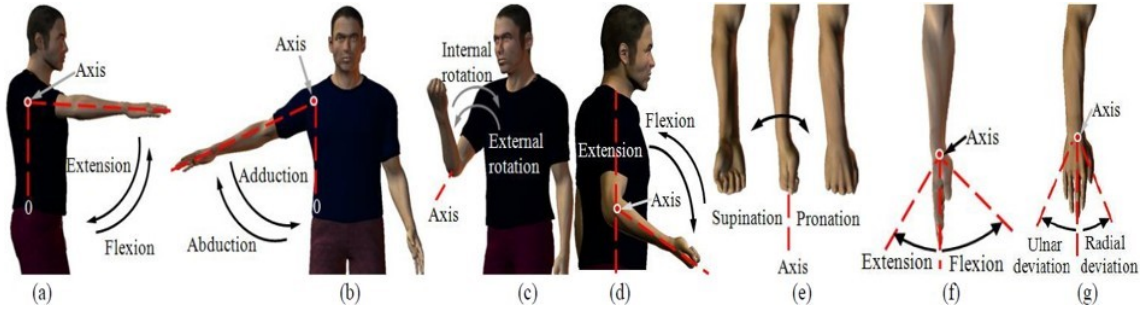


Fig. 1. Motions of the human arm.

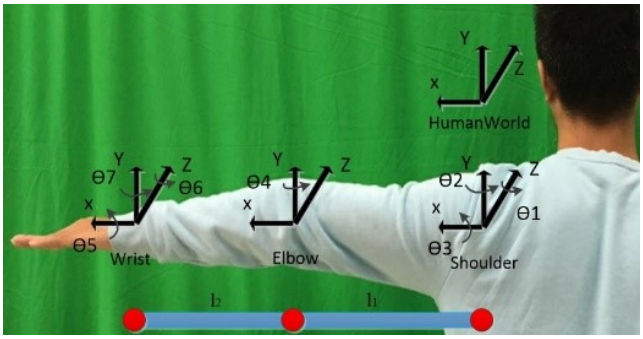


Fig. 2. Frames on the human arm.

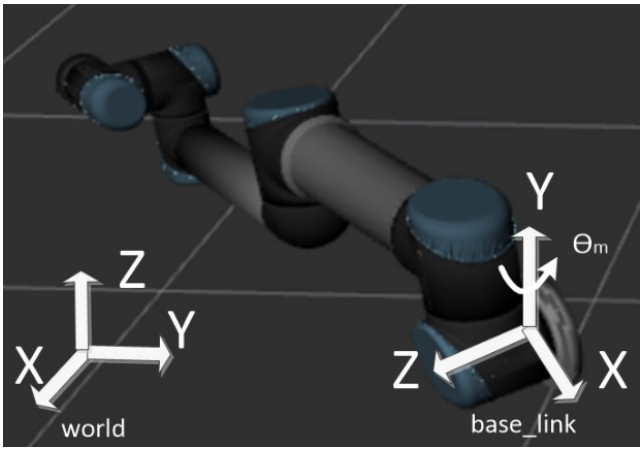


Fig. 3. Initial configuration.

tracks the direction of the human upper-arm, v_{link3} tracks the direction of the human forearm and the last three links of the UR5 robot perform a bit like a trans-formative hand, moving with the rotation of the human wrist. To achieve the above performance, three metrics are employed in the following part.

For the first virtual joint pair, the metric used is presented as follows

$$M_1 = \arccos\left(\frac{v_{link2} \cdot v_{upper-arm}}{|v_{link2}||v_{upper-arm}|}\right) \tag{3}$$

where $arccos()$ is the arc-cosine function that ranges from 0 to π.

For the second virtual joint pair, the joint-level metric is used

$$M_2 = ||\theta_e - \theta_4^D||_2 \tag{4}$$

$$\theta_4^D = \arccos\left(\frac{v_{upper-arm} \cdot v_{forearm}}{|v_{upper-arm}||v_{forearm}|}\right) \tag{5}$$

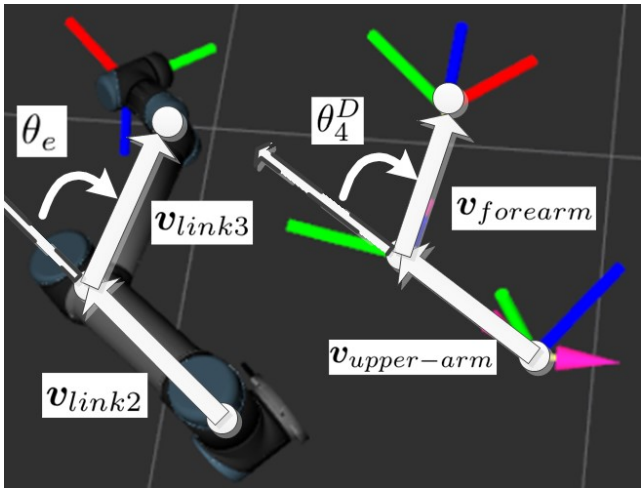


Fig. 4. Illustration of the metrics used in the first and second virtual joint pairs.

For the third virtual joint pair, the rotation of the frame6 with respect to the frame3 of the UR5 robot is expected to be equal to the rotation of the wrist frame with respect to elbow frame (as shown in Fig. 5). The metric used is presented as

$$M_3 = ||R_c R_{Wrist}^{Elbow} (R_c)^{-1} - R_{Frame6}^{Frame3}||_2 \tag{6}$$

where R_{Wrist}^{Elbow} represents the orientation of the wrist frame with respect to the elbow frame, R_{Frame6}^{Frame3} represents the orientation of the frame6 of the UR5 robot with respect to its frame3, and

$$R_c = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \tag{7}$$

4.2 Computational Procedure

(1) For virtual joint pair1 and the determination of θ_m

The goal of this step is to obtain the desired joint angles θ_1^I and θ_2^I by minimizing the metric (3) and to determine θ_m.

A new frame, Joint1-frame, whose direction always parallel to those of the base-link-frame, is attached to the origin of the frame1. Then in the initial configuration of the UR5 robot, the locations of these points in the Joint1-frame are shown in Fig. 6(a). The desired joint angles θ_1^I and θ_2^I can be obtained as follows:

$$\theta_1^I = {}^0\theta_1^I - \angle AOC = -\angle AOC \tag{8}$$

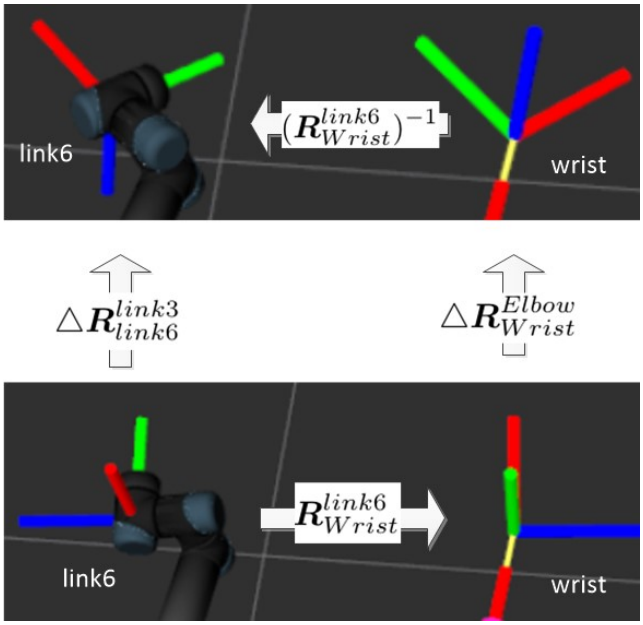
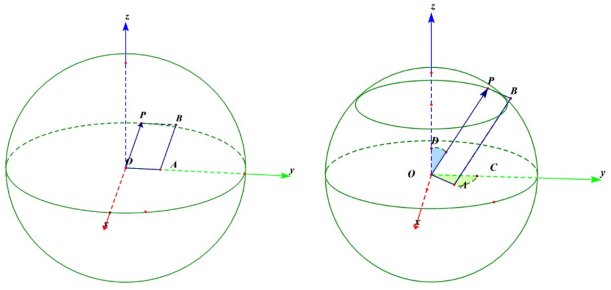


Fig. 5. Illustration of the metric used in the third virtual joint pair.

$$\theta_2^I = {}^0\theta_2^I + \frac{\pi}{2} - \angle POD = \frac{3\pi}{2} - \angle POD \tag{9}$$

where $\angle POD = \arccos\left(\frac{z\cdot\overrightarrow{OP}}{|z||\overrightarrow{OP}|}\right)$, $\angle AOC = \arccos\left(\frac{y\cdot\overrightarrow{OA}}{|y||\overrightarrow{OA}|}\right)$, $\overrightarrow{OA} = \overrightarrow{OP} \times z$, $\overrightarrow{OP} = v_{link2}$, $z = [0\ 0\ 1]^T$, $y = [0\ 1\ 0]^T$.



(a) (b)

Fig. 6. Illustration of the geometric relationship for first virtual joint pair.

Though the solutions (8)-(9) can bring the metric (3) to zero, it cannot reproduce the shoulder internal/external rotation in the UR5 robot. The underlying cause lies in one less DOF in the first virtual joint of the UR5 robot than that of the human arm. Ignoring the shoulder internal/external rotation will lead to remarkable tracking errors in the location of the human wrist. To solve this problem, (1) the human operator is suggested to avoid the shoulder internal/external rotation; (2) a proper θ_m should be chosen to make the rotation axis of the elbow flexion/extension motion parallel to that of the joint 3 of UR5 as nearly as possible when the human arm moves in its work-space. We will show how to determine θ_m in the following parts.

According to the solutions (8)-(9), the rotation axis of the elbow flexion/extension would parallel to that of the joint 3 of the UR5 when v_{link2} moves in the x-y plane of the Joint1 frame (as shown in Fig. 7(a)), while those two axes would be vertical to each other when v_{link2} moves in the y-z plane (as shown in Fig. 7(b)). Consequently, it is preferable that v_{link2}, which follows the direction of $v_{upper-arm}$, moves near the x-y plane of Joint1 frame when the human arm moves in its workspace. As a result, θ_m is chosen as $\pi/4$ according to the main range of human arms motion, as shown in Fig. 7(c).

(2) For virtual joint pair2. The desired joint angle θ_3^I can be obtained directly with the following equation,

$$\theta_3^I = {}^0\theta_3^I - \theta_e = -\theta_4^D \tag{10}$$

(3) For virtual joint pair3. According to rotation order of the last three joints of the UR5 robot, R_{Wrist}^{Elbow} has to be converted to the intrinsic X-Y-Z Euler angle $(\alpha_W, \beta_W, \gamma_W)$ so that it can be executed by the UR5 robot. Then the desired joint angles θ_4^I , θ_5^I and θ_6^I can be obtained as:

$$\theta_4^I = {}^0\theta_4^I + \alpha_W = \alpha_W \tag{11}$$

$$\theta_5^I = {}^0\theta_5^I + \beta_W = \pi/2 + \beta_W \tag{12}$$

$$\theta_6^I = {}^0\theta_6^I + \gamma_W = \gamma_W \tag{13}$$

5. EXPERIMENTS

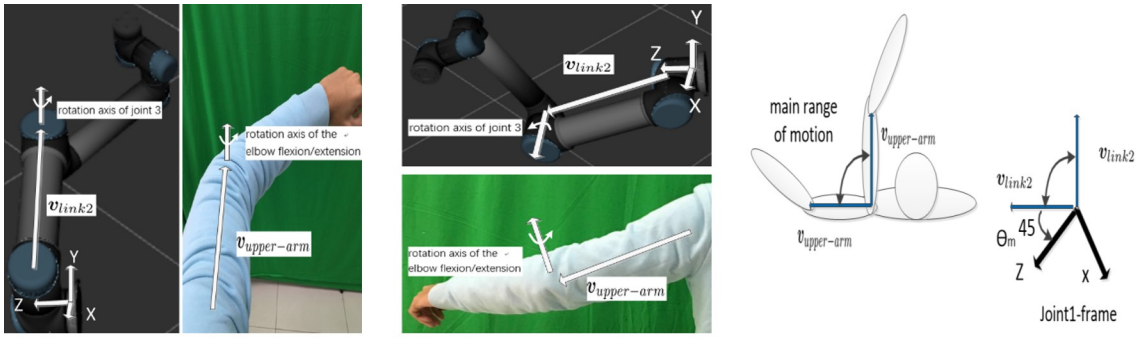
5.1 Experiment settings

System composition. The imitation system mainly consists of the human operator, the Perception Neuron motion capture system, kinematics mapping module and the UR5 robot, as is shown in Fig. 8. Communications among the last three modules are realized through ROS (Robot Operating System) framework.

Parameters settings. The Perception Neuron motion capture system maps the human arms motion to the motion of an avatar and generates the BVH data files. In the experiment, the length of the upper-arm of the avatar is set as $l_1 = 0.7\,|a_2| = 297.5$mm, and the length of the forearm is set as $l_2 = 0.7\,|a_3| = 274.6$mm. The velocity limit is set to 3.15 rad/s for every joint of the robot. The target states for the UR5 robot that violate the joint velocity limit would be abandoned.

System input and collected data. The motion of human arms can be roughly divided into two modes, i.e., position-changing mode and orientation-changing mode. In the position-changing mode, the operator rotates his first and second virtual joint, consequently changing the location of the upper-arm and forearm, as well as the position of the Wrist-frame. In the orientation-changing mode, the operator rotates his third virtual joint, consequently changing the orientation of the wrist frame. Two demos of the human arm's motion are recorded via the Perception Neuron system and then sent to the kinematics mapping module. To show the effectiveness of the proposed method, Cartesian-space-based mapping (CSBM) approach with workspace-matching and Joint-space-based mapping (JSBM) approach are also implemented to compare with the developed scheme.

In demo♯1, the human arm mainly moves in the position-changing mode; while in demo♯2, the human arm mainly



(a) (b) (c)

Fig. 7. Illustration of the determination of θ_m.

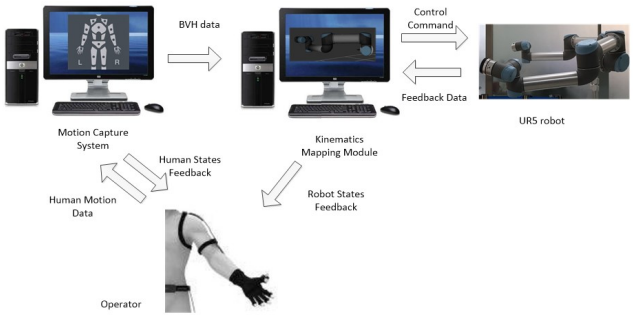


Fig. 8. Components of the imitation system.

moves in the orientation-changing mode. Both demos are used as inputs to the proposed approach, the Cartesian-space-based mapping approach and the joint-space-based mapping approach. To separate the imitation errors, the desired robot motions are obtained by computing the forward kinematics for the desired joint states, instead of the real UR5's motion data, which are then collected and compared with human motion data obtained by the motion capture system (as shown in Fig. 9).

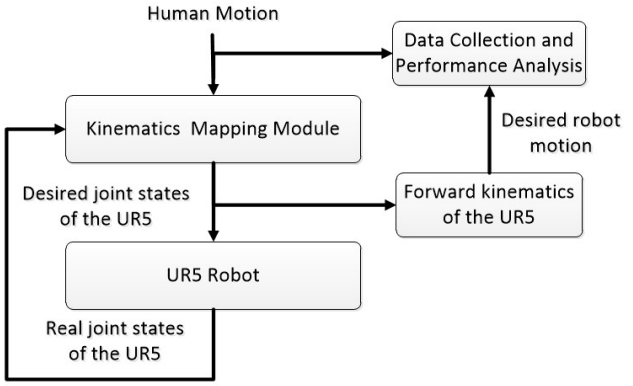


Fig. 9. Data flow of the performance analysis procedure.

The following statistics are employed to evaluate the imitation performance.

a) Similarity in the end-effector locations.

It is important to measure the similarity between end-effector location of the demonstrator and that of the imitator. The Cartesian coordinate of $v_{Wrist}^{Shoulder}$ is considered as the tip position of human arm and recorded as (x^D, y^D, z^D). The orientation of Wrist frame with respect to the world frame is considered as the tip orientation of human arm and recorded as RPY angles. As for the tip position of the UR5 robot (x^I, y^I, z^I) in the proposed approach, it refers to the Cartesian coordinate of the vector from the origin of the frame1 to that of the frame3; while in the CSBM it refers to the coordinate of the vector from the origin of the base link-frame to that of the frame6. In both approaches, the tip orientation of the UR5 robot $(\alpha^I, \beta^I, \gamma^I)$ refers to the orientation of the frame6 of the UR5 with respect to the World frame, transformed into the RPY angles. The root-mean-square errors (RMSE) between the relative movement of the tip of the human arm and that of the UR5 robot can be computed as follows

$$\begin{cases} RMSE\_d\xi = \sqrt{\dfrac{1}{n-1}\sum_{i=1}^{n-1}\left(d\xi^D(i) - 0.7d\xi^I(i)\right)^2} \\ d\xi^D(i) = \xi^D(i+1) - \xi^D(i) \\ d\xi^I(i) = \xi^I(i+1) - \xi^I(i) \end{cases} \tag{14}$$

where $\xi = x, y, z, \alpha, \beta, \gamma$.

b) Accuracy of algorithms.

In the experiment, $err_{upperarm}$ and $err_{forearm}$ are employed to measure the similarity in the configuration of the human arm and UR5 robot. The following statistics are computed:

$$\begin{cases} RMSE\_upperarm = \sqrt{\dfrac{1}{n}\sum_{i=1}^{n}\left(err_{upperarm}(i)\right)^2} \\ err_{upperarm}(i) = \arccos\left(\dfrac{v_{link2}(i)\cdot v_{upper-arm}}{|v_{link2}||v_{upper-arm}(i)|}\right) \end{cases} \tag{15}$$

$$\begin{cases} RMSE\_forearm = \sqrt{\dfrac{1}{n}\sum_{i=1}^{n}\left(err_{forearm}(i)\right)^2} \\ err_{forearm}(i) = \arccos\left(\dfrac{v_{link3}(i)\cdot v_{forearm}(i)}{|v_{link3}(i)||v_{forearm}(i)|}\right) \end{cases} \tag{16}$$

c) Smoothness of imitation motion.

The following statistic is used to measure the smoothness of the imitation motion.

$$\begin{cases} MEAN\_SumJoints = \dfrac{1}{n-1}\sum_{i=1}^{n-1}\sum_{j=1}^{6}|d\theta_j^I(i)| \\ d\theta_j^I(i) = \theta_j^I(i+1) - \theta_j^I(i) \end{cases} \quad (17)$$

d) Time efficiency of algorithms.

The consuming time of kinematics mapping, which consists in the time for the transform data and that for the computation of the desired joint angles for the UR5 robot, is recorded with the help of the clock( ) function provided by the C Time library. The average of the consuming time is computed as below

$$MEAN\_time = \frac{1}{n}\sum_{i=1}^{n} t(i) \quad (18)$$

where $t(i)$ is the consuming time of the ith kinematics mapping procedure.

5.2 Experiment results

Results of the proposed approach and CSBM. The statistical data is shown in Table 1, where Approach♯1 refers to the proposed mapping approach. It can be seen that, with the proposed mapping approach, the frame3 of the UR5 robot tracks the displacement of the Wrist-frame successfully in both demos, and the position tracking errors along the Z axis are a litter larger than those along the X and Y axes. The orientation of frame6 imitates that of the Wrist-frame at a high accuracy. The $RMSE\_d\alpha$, $RMSE\_d\beta$, $RMSE\_d\gamma$ are all less than 1°. $err_{upperarm}$ is close to zero all the time. $err_{forearm}$ is about 5° in the demos. It means that the configuration of the UR5 robot tracks that of the human arm nearly perfectly (as shown in Fig.15). The $MEAN\_SumJoints$ is about 4° in the first demo and 3° in the second demo. The consuming time for kinematics mapping is less than 0.2ms in most cases. The $MEAN\_time$ is about 0.1ms. It is worth mentioning that the advantage of the proposed method is the capability of imitation simultaneously in Cartesian space and joint space, not only about the accuracy of end-effector tracking, as shown in Fig. 10 and Table 1.

With the CSBM approach, the frame6 of the robot tracks the position and orientation of Forearm-frame perfectly. $err_{upperarm}$ and $err_{forearm}$ are much larger than those in the virtual-joint-based approach. The $RMSE\_err_{upperarm}$ is about 10° and the $RMSE\_err_{forearm}$ is about 23°. It means that there are some differences between the configuration of the UR5 robot and the human arm. In the first demo, the $MEAN\_SumJoints$ in the two approaches are pretty close. But in the second demo, the $MEAN_SumJoints$is 13% larger in CSBM approach. Also, the average of consuming time of the CSBM is about 50% larger than that in the proposed approach, though they are both less than 0.2ms.

Results of JSBM approach. Fig. 10 shows the mapping results of the proposed approach and the JSBM approach in the second demo. It can be known that, when the human operator uses only one of the last three DOFs of his arm, i.e., changes the orientation of his wrist frame along one of its axes (as shown in Fig. 10. (a), (d), (g)), both the proposed approach and the JSBM approach have satisfying tracking performance in wrist rotation. However, when composite motions are taken, such as doing elbow flexion/extension and wrist flexion/extension simultaneously (as shown in Fig.10. (j)), the proposed approach tracks the rotation of wrist frame successfully while the JSBM fails to guarantee it.

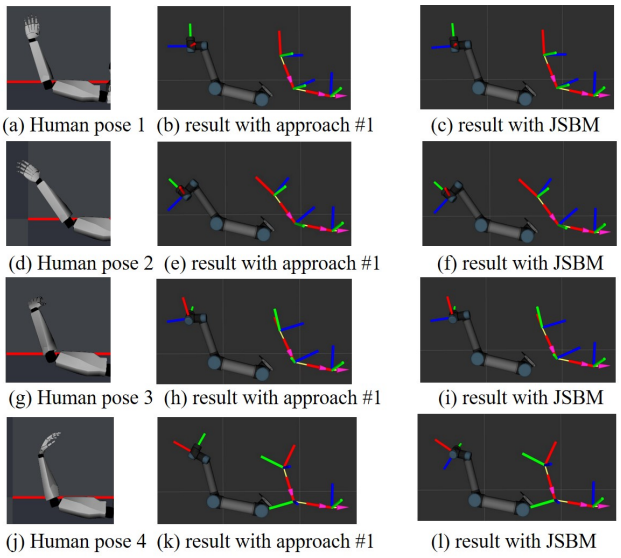


Fig. 10. Mapping results with the proposed approach and the JSBM.

5.3 Discussions

Effects of experiment conditions. The initial pose of the UR5 robot has some important effects on the kinematics mapping procedure. To be specific, for the mapping in the virtual joint pair♯3, the initial pose of the robot decides which type of Euler-angle the rotation matrix R_{Wrist}^{Elbow} has to be converted to. It also influences the computation procedure for the virtual joint pair♯1.

In the proposed approach and CSBM approach, the tuning of the lengths of the avatar's upper-arm and forearm and the scaling factor k help to match the workspace of the UR5 robot to that of the human arm. The motions of human arms generally fall into two categories: the position-changing and the orientation-changing modes. We have studied the imitation performance of our system with two types of human motions as input respectively.

Table 1. Experiment results

Statistics	Approach♯1 in demo♯1	CSBM in demo♯1	Approach♯1 in demo♯2	CSBM in demo♯2
$R\_dx(mm)$	0.3117	$5.15e{-4}$	0.0522	$5.18e{-4}$
$R\_dy(mm)$	0.5468	$4.20e{-4}$	0.2123	$5.02e{-4}$
$R\_dz(mm)$	2.3279	$3.22e{-4}$	1.1346	$5.24e{-4}$
$R\_d\alpha(\circ)$	0.4171	$3.84e{-4}$	0.2666	$3.07e{-4}$
$R\_d\beta(\circ)$	0.6728	0	0.2011	0
$R\_d\gamma(\circ)$	0.3796	0	0.1527	0
$R\_upperarm(\circ)$	7.17e-5	12.7535	6.91e-5	8.2251
$R\_forearm(\circ)$	3.5605	23.4819	6.2294	22.1848
$M\_SumJoints(\circ)$	4.1739	4.2360	2.9939	3.4019
$M\_time(ms)$	0.0966	0.1502	0.1144	0.1561

* R and M represent the abbreviation of RMSE and MEAN.

Compared with the CSBM approach. From the experiment results it can be known that both approaches have a good performance in imitating the location of the human wrist. The location tracking errors of CSBM are mainly caused by calculation errors. In the proposed approach, the location tracking errors are resulted mainly from the dissimilar embodiments in the first virtual joint pair. To be specific, the first virtual joint of the human arm has one more DOF than that of the UR5 robot, and their morphologies are different. The relatively larger position tracking errors along the Z direction is an evidence for this. The proposed approach does a much better job in imitating the configuration. In the proposed approach, the configuration of the UR5 robot is almost the same as that of the human arm. But in CSBM, the similarity in the configuration of the UR5 and that of the human arm can't be guaranteed, even the size parameters of the avatar in motion capture system are well tuned and a proper initial configuration is chosen. In the first demo, two approaches have similar performance in the smoothness of robot motion; while in the second demo, the proposed approach does a better job than the Cartesian space-based approach in terms of the motion smoothness. In the second demo, the human operator mainly moves his third virtual joint of arm, corresponding to the third virtual joint of the UR5 robot in the proposed approach. So in the proposed approach, mainly the last three joints of the robot have to move to imitate the human arm's motion in the second demo while all six joints have to be used in CSBM.

Also, the proposed approach has a higher computation efficiency. In the proposed approach, the overall kinematics mapping problem is decomposed into finding mapping solutions for three virtual joint pairs, so that the inverse-kinematics for a 6-DOF structure is avoid. In CSBM, the algorithm is consisted of three parts–listening to the transform data, calculating all feasible inverse-kinematics solutions and choosing a solution closest to the current states as the desired joint angles for UR5.

To sum up, the proposed approach has better performance on imitating the configuration of the human arm, better motion smoothness and higher computation efficiency, at the price of the less satisfying performance in tracking the location of the human wrist. In those where the similarities in the tip location and configuration are both required, the proposed approach has advantages over CSBM.

Compared with JSBM. By properly choosing initial poses for the UR5 robot and the human arm and designing a suitable JSBM, the UR5 robot would track the rotation of the human Wrist-frame successfully if the Wrist-frame only rotates along one of its axes. However, when the human operator uses more than one of his arm's last three DOFs simultaneously, JSBM would fail to track. It is one of the limitations for applying JSBM to kinematics mapping across dissimilar embodiments. In some respects, JSBM tries to find the mapping that maps the effect of rotation (or displacement) of each DOF of the demonstrator to that of the corresponding DOF of the imitator. However, in cases where the demonstrator and the imitator have different topologies, mapping at the DOF-level can't guarantee satisfying overall performance. To overcome this challenge, the proposed approach divides the DOFs of the demonstrator and those of the imitator into the same number of groups (called as virtual joints) and tries to find the virtual-joint-level mapping that maps the effect of each virtual joint of the demonstrator to that of the corresponding one of the imitator. By choosing different grouping schemes and describing different aspects of effects of the virtual joints, diverse virtual-joint-based mapping strategies can be defined. So the proposed approach is much more flexible and powerful than JSBM and can be applied to kinematics mapping across embodiments that have different DOFs and dissimilar topologies.

6. CONCLUSION

In this paper, an on-line imitation system for the human arm and an UR5 robot has been developed, in which the tip location and configuration of the human arm can be imitated simultaneously by the robot. A virtual-joint-based approach has been proposed to tackle the kinematics mapping between dissimilar embodiments. Experiments have shown that, compared with the traditional Cartesian-space-based approach, the proposed approach has better performance in human arm configuration imitating, better motion smoothness and higher computation efficiency. Also, the intuitive correspondence between the virtual joints of the human arm and that of the UR5 would enable the human operator to control the robot in a more natural way. On the other hand, compared with simple and intuitive joint-space-based mapping approach, the proposed approach is more flexible and can be applied to kinematics mapping with different DOFs and dissimilar topologies.

REFERENCES

Alissandrakis, A., Nehaniv, C.L., and Dautenhahn, K. (2007). Correspondence mapping induced state and action metrics for robotic imitation. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 37(2), 299–307.

Argall, B.D., Chernova, S., Veloso, M., and Browning, B. (2009). A survey of robot learning from demonstration. *Robotics and Autonomous Systems*, 57(5), 469–483.

Calinon, S. and Billard, A. (2007). Incremental learning of gestures by imitation in a humanoid robot. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, 255–262. ACM.

Chen, Z., Wang, Z., Liang, R., Liang, B., and Zhang, T. (2020). Virtual-joint based motion similarity criteria for human–robot kinematics mapping. *Robotics and Autonomous Systems*, 125, 103412.

Gopura, R. and Kiguchi, K. (2009). Mechanical designs of active upper-limb exoskeleton robots: State-of-the-art and design difficulties. In *2009 IEEE International Conference on Rehabilitation Robotics*, 178–187.

Hwang, C.L., Chen, B.L., Syu, H.T., Wang, C.K., and Karkoub, M. (2014). Humanoid robot's visual imitation of 3-d motion of a human subject using neural-network-based inverse kinematics. *IEEE Systems Journal*, 10(2), 685–696.

Jha, A., Chiddarwar, S.S., Bhute, R.Y., Alakshendra, V., Nikhade, G., and Khandekar, P.M. (2017). Imitation learning in industrial robots: A kinematics based trajectory generation framework. In *In Proceedings of the Advances in Robotics*, 1–6.

Kormushev, P., Nenchev, D.N., Calinon, S., and Caldwell, D.G. (2011). Upper-body kinesthetic teaching of a free-standing humanoid robot. In *2011 IEEE International Conference on Robotics and Automation*, 3970–3975. IEEE.

Kufieta, K. (2014). Force estimation in robotic manipulators: Modeling, simulation and experiments. *Department of Engineering Cybernetics NTNU Norwegian University of Science and Technology*.

Laguillaumie, P., Laribi, M.A., Seguin, P., Vulliez, P., Decatoire, A., and Zeghloul, S. (2016). From human motion capture to industrial robot imitation. In S. Zeghloul, M.A. Laribi, and J.P. Gazeau (eds.), *Robotics and Mechatronics*, 301–312. Springer International Publishing, Cham.

Lee, J.H. et al. (2012). Full-body imitation of human motions with kinect and heterogeneous kinematic structure of humanoid robot. In *2012 IEEE/SICE International Symposium on System Integration (SII)*, 93–98. IEEE.

Li, C., Yang, C., Wan, J., Annamalai, A., and Cangelosi, A. (2017). Neural learning and kalman filtering enhanced teaching by demonstration for a baxter robot. In *2017 23rd International Conference on Automation and Computing (ICAC)*, 1–6.

Lin, H.I., Liu, Y.C., and Lin, Y.H. (2014). Intuitive kinematic control of a robot arm via human motion. *Procedia Engineering*, 79, 411–416.

Nehaniv, C.L. and Dautenhahn, K. (2000). Of hummingbirds and helicopters: An algebraic framework for interdisciplinary studies of imitation and its applications. In *Interdisciplinary approaches to robot learning*, 136–161. World Scientific.

Shahverdi, P. and Masouleh, M.T. (2016). A simple and fast geometric kinematic solution for imitation of human arms by a nao humanoid robot. In *2016 4th International Conference on Robotics and Mechatronics (ICROM)*, 572–577. IEEE.

Suleiman, W., Yoshida, E., Kanehiro, F., Laumond, J.P., and Monin, A. (2008). On human motion imitation by humanoid robot. In *2008 IEEE International Conference on Robotics and Automation*, 2697–2704. IEEE.

Wang, Z., Liang, B., Sun, Y., and Zhang, T. (2020). Adaptive fault-tolerant prescribed-time control for teleoperation systems with position error constraints. *IEEE Transactions on Industrial Informatics*, 16(7), 4889–4899.

Wang, Z., Chen, Z., Zhang, Y., Yu, X., Wang, X., and Liang, B. (2019a). Adaptive finite-time control for bilateral teleoperation systems with jittering time delays. *International Journal of Robust and Nonlinear Control*, 29(4), 1007–1030.

Wang, Z., Sun, Y., and Liang, B. (2019b). Synchronization control for bilateral teleoperation system with position error constraints: A fixed-time approach. *ISA transactions*, 93, 125–136.

Xiaojun Zhao, Qiang Huang, Peng Du, Dongming Wen, and Kejie Li (2004). Humanoid kinematics mapping and similarity evaluation based on human motion capture. In *International Conference on Information Acquisition, 2004. Proceedings.*, 426–431.

Zhang, L., Cheng, Z., Gan, Y., Zhu, G., Shen, P., and Song, J. (2016). Fast human whole body motion imitation algorithm for humanoid robots. In *2016 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 1430–1435. IEEE.

Zhu, Z. and Hu, H. (2018). Robot learning from demonstration in robotic assembly: A survey. *Robotics*, 7(2), 17.