DOI: 10.19650/j.cnki.cjsi.J2108032

基于强化学习的机器人手臂仿人运动规划方法*

杨傲雷[1,2],陈燕玲[1],徐昱琳[1]

(1. 上海大学机电工程与自动化学院 上海 200444; 2. 上海市电站自动化技术重点实验室 上海 200444)

摘 要:面向人-机器人交互共融环境对机械臂仿人运动规划的重大需求,本文提出了一种基于强化学习的机器人手臂仿人运动规划方法。首先,基于人体手臂的结构特征,设计了体现机械臂运动特性的肩夹角、肘夹角和腕关节运动角,并采用正态性和相关性分析方法,对 VICON 运动捕捉系统获取的人体手臂运动数据进行分析,以获取人臂运动特性规则。然后,根据不同的运动特性规则,设计对应的回报函数,并采用强化学习方法进行机械臂仿人运动模型的训练。最后,搭建机械臂仿人运动平台,实验统计仿人运动的成功率为 91.25%,验证了所提规划方法的可行性和有效性,可用于提高机械臂运动的仿人性。

关键词: 人臂运动特性;仿人运动规划;强化学习;运动捕捉系统

中图分类号: TP391 TH86 **文献标识码**: A **国家标准学科分类代码**: 510.4050

Humanoid motion planning of robotic arm based on reinforcement learning

Yang Aolei[1,2], Chen Yanling[1], Xu Yulin[1]

(*1. School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444, China;*
2. Shanghai Key Laboratory of Power Station Automation Technology, Shanghai 200444, China)

Abstract: To meet the requirement of humanoid motion planning of the robotic arm in human-robot interaction environment, a humanoid motion planning method of the robotic arm based on reinforcement learning is proposed in this article. Firstly, based on the structural characteristics of the human arm, the shoulder angle, the elbow angle and the wrist joint motion angle are designed to reflect motion characteristics of the robotic arm. The motion data of human arm captured by the VICON system are analyzed by using normality and correlation analysis methods to achieve the motion characteristics of the human arm. Then, according to different motion characteristics rules, the corresponding reward functions are designed, and the humanoid motion model is trained by the reinforcement learning method. Finally, the humanoid motion platform of the robot arm is established, and the success rate of the humanoid motion is 91.25%. It evaluates the feasibility and effectiveness of the proposed method, which could be used to improve the humanization of robot motion.

Keywords: human arm motion characteristics; humanoid motion planning; reinforcement learning; motion capture system

0 引 言

随着机器人与人工智能技术的发展,机器人的设计开始转向能够实现人类与机器人之间的自然交互[1]。为了使人类能够对机械臂的运动做出友好的预判和理解,保证人和机械臂在人机交互中的安全,不仅要求机械臂在结构上与人体手臂相似,更要求其在运动上具备人体手臂运动的特点[2]。机械臂仿人运动的规划主要是从人体手臂出发,分析不同动作状态下的运动特征,并将其应用到机械臂中。因此,针对人体手臂运动的分析是仿人运动的关键。

目前,国内外研究学者已提出多种机械臂仿人运动规划方法。常见的方法是通过研究采集的运动数据,分析人体手臂运动过程的特征,并利用这些特征完成机械臂的仿人规划。赵京等[3]主要通过最小化机械臂运动过

收稿日期:2021-06-02 Received Date: 2021-06-02
*基金项目:国家自然科学基金(61873158)、上海市自然科学基金(18ZR1415100)项目资助

程中的势能来引导仿人轨迹的生成。陈盛等[4]则利用最小势能进一步确定机械臂肘部位置以实现仿人运动规划。Li 等[5]通过最小化肌肉强度确定机械臂的拟人姿态,并将其作为生成仿人轨迹的标准。为了进一步研究机械臂的仿人运动,研究者们还结合人体手臂的运动,设计仿人变量、建立仿人标准和约束。Gong 等[6]则根据人体手臂的结构和运动方式,通过 4 个变量定义了不同的人臂运动原语。García 等[7]和 Rosell 等[8]先后通过设计吸引因子和不同的抓握类型以引导机械臂生成仿人路径。参考人臂运动过程中的阻抗特征,Yang 等[9]设计了人臂末端刚度。这类方法虽然能够独立的生成仿人运动轨迹,但大多是针对某一类特定任务的仿人运动规划,同时,仿人变量的设计较为单一,仿人参数的选定容易对仿人运动造成影响。

随着智能学习方法的发展,机械臂仿人运动的研究得到了进一步的拓展。Zhu 等[10]使用遗传算法,并设计共享控制角度以控制机械臂实现仿人运动。Wei 等[11]则通过已知的机械臂末端位置方向,设计机械臂末端最佳姿态,并通过优化 5 个目标函数以实现机械臂的仿人行为。Duarte 等[12]通过设计不同的运动状态,并采用高斯混合模型完成模型的计算,实现对机器人仿人运动的控制。虽然结合机器学习的仿人规划方法,可以提升机械臂仿人运动的准确性和效率,但它存在训练或示教数据不足、数据模型无法覆盖人体手臂运动模型的情况,影响机械臂仿人运动的鲁棒性。

强化学习的运动规划不仅能够有效地避免人为设置仿人参数的不合理性,还能规避训练数据不足的问题。因此,本文提出了一种基于强化学习的机器人手臂仿人运动规划方法。首先,结合人体手臂运动范围,采用 VICON 系统获取人臂达点运动数据以保证数据的准确性;其次,与单一的仿人变量不同,本文创新性地设计了表征人体手臂基本运动特征的肩夹角、肘夹角和腕关节运动角,并通过正态性和相关性分析方法获取人体手臂动作特性规则,它们能够体现人体手臂各关节的运动关系;然后,根据不同规则,设计对应的仿人回报函数,采用仿人规划(humanoid planning, HP)强化学习算法进行机械臂的仿人运动训练;最后,通过大量的机械臂达点运动实验,验证所提方法的仿人性、可行性和有效性。

1　问题描述及方法架构

1.1　人体手臂与机械臂之间的映射关系

将人体手臂运动特征融合到机械臂中是实现机械臂仿人运动的关键,这就要求获取人体手臂与机械臂之间的映射关系。本文采用 Kinova 公司的 Jaco2 机械臂,该机械臂共有 6 个自由度。参考人体手臂与 Jaco2 各关节的运动方式以及运动时各关节角的定义,给出图 1 所示的人体手臂与 Jaco2 在矢状面上的映射关系。图中人体手臂浅色竖直向上的姿态对应机械臂浅色竖直向上的姿态。当人体手臂肩关节运动了 $\boldsymbol{\theta}_{hs}$ rad 时,机械臂肩关节对应的角度为 $\boldsymbol{\theta}_{js}=(\pi-\boldsymbol{\theta}_{hs})+\pi$。当人体手臂肘关节运动了 $\boldsymbol{\theta}_{he}$ rad 时,机械臂肘关节对应的角度为 $\boldsymbol{\theta}_{je}=\pi-\boldsymbol{\theta}_{he}$。

根据文献[13-14]中人体手臂各关节的运动范围和机械臂的运动范围,通过上述公式能够计算出机械臂肩关节和肘关节的仿人运动范围,即机械臂关节 2 的仿人运动范围为[3.438,5.76] rad,关节 3 的仿人运动范围为[0.559,π] rad。由于 Jaco2 机械臂的关节 1 和腕关节(即关节 4~6)分别对应于人体自身的旋转行为和人体手臂腕关节的环转运动,因此机械臂的关节 1 和腕关节的运动范围不受影响,均为[-2π,2π] rad。

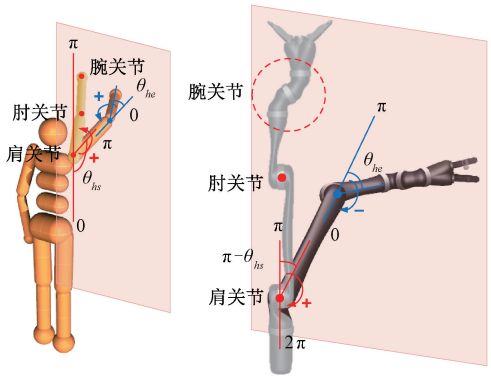


图 1　人体手臂与 Jaco2 机械臂之间的映射

Fig. 1　Mapping between human arm and Jaco2

1.2　整体方法架构

本文所提方法的架构体系如图 2 所示,主要包含人臂关节变量提取、回报函数构建和仿人规划强化学习。人臂关节变量提取主要通过 VICON 系统采集大量的人体手臂运动数据,并计算对应的人臂关节变量。经过正态性分析和相关性分析,提取相应的特征,包括肩夹角 α_1 和肘夹角 α_2 的运动范围,以及腕关节运动角 β_1 的运动趋势和收敛值。此部分的作用是为后续强化学习训练部分提供机械臂仿人运动的“先验知识”。回报函数构建部分主要是融合前面的运动规律,设计对应的回报函数,简化仿人运动的训练。另外,仿人规划强化学习算法结合了 DDPG(deep deterministic policy gradient[15])和 HER(hindsight experience replay[16])算法,用于进行机械臂仿人运动的训练学习。其中,DDPG 算法主要用于实现对连续动作的训练,HER 算法则主要通过利用失败的探索结果,加快 DDPG 算法的训练速度。最后,采用 TCP/IP 通信的方式将训练生成的仿人运动路径实时传输给真实的机械臂仿人运动平台。

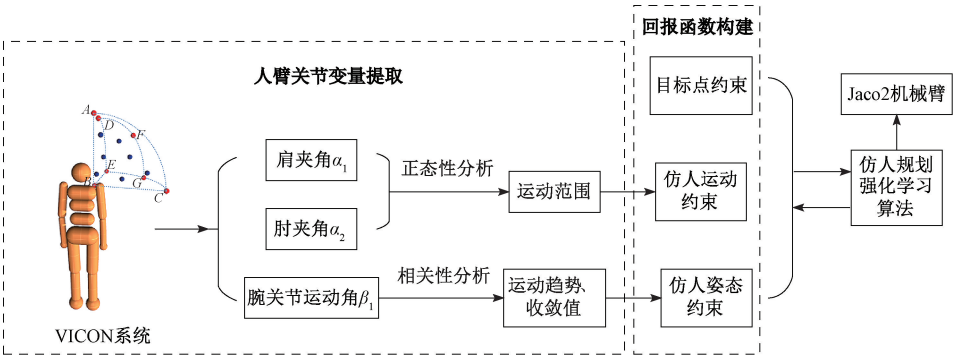


图 2 整体方法架构

Fig. 2 Architecture of the proposed method

2 手臂运动规律分析与学习

本文采用 VICON 系统采集人体手臂达点运动(即到达目标点运动)的数据,通过对数据的分析与学习,获取手臂在达点运动过程中的典型规律和约束。

2.1 目标点的选择

对于空间中任意可达位置的目标点,人们通常能够通过旋转自身的方式实现达点运动。对应的,Jaco2 机械臂可通过旋转关节 1 的方式将原目标点转换到机械臂的右半空间中。此外,由于本文研究的这类机械臂的运动空间在基座以上,因此在进行人体右臂实验时,只需针对人体右臂右半空间中肩关节以上的位置进行达点实验即可,并将该运动空间定义为 $\boldsymbol{S}_M$。

为了研究人体手臂在 $\boldsymbol{S}_M$ 中的运动特征,本文对图 3 中的 14 个目标点进行达点运动实验。其中,点 A、B、C、D、E、F、G 为空间 $\boldsymbol{S}_M$ 的边界点,其他未标注的点为空间 $\boldsymbol{S}_M$ 内的随机点。参考人体手臂各关节的运动范围,能够确定 $\boldsymbol{S}_M$ 的边界位置。表 1 所示为 $\boldsymbol{S}_M$ 中各边界点的定义。

2.2 基于 VICON 的人体手臂运动数据采集方法

VICON 系统是一种光学动作捕捉系统,具备实时、精准、稳定的特点。基于 VICON 系统的人臂运动数据采集,首先需要设计合适的刚体标志块,以获取运动过程中的人臂空间位置信息。因为人体手臂由肩、上臂和前臂 3 部分组成,所以本文设计了 3 个刚体标志块,即肩标志块、上臂标志块和前臂标志块,它们分别由 4 个排布方式完全不同的反光球构建,保证了各标志块之间的区别,避免了 VICON 系统的错误识别,如图 4(a)~(c)所示。肩关节由肩标志块中的 $S4$ 表示,肩向量 $\vec{\boldsymbol{V}}_S$ 由 $S1$ 和 $S4$ 构建;上臂向量$\vec{\boldsymbol{V}}_A$ 由上臂标志块中的 $A1$ 和 $A4$ 构建;肘关节由前臂标志块中的 $F1$ 表示,腕关节由 $F4$ 表示,前臂向量 $\vec{\boldsymbol{V}}_F$ 由 $F1$ 和 $F4$ 构建。

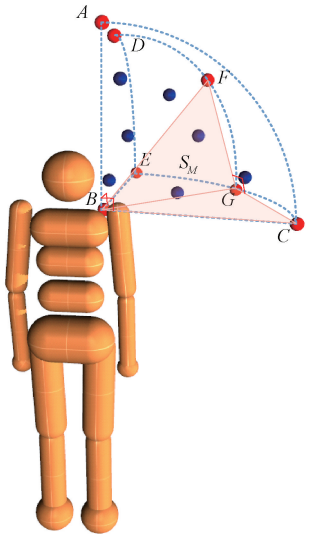


图 3 人体右臂运动空间及目标点的选取

Fig. 3 Motion space of human right arm and selection of target points

表 1 人体右臂运动空间中各边界点的定义

Table 1 Boundary points in motion space of human right arm

边界点	肩关节运动角度/rad	肘关节运动角度/rad
A	冠状面:外展 π	矢状面:前屈 0
B	不动	矢状面:前屈 2.496
C	冠状面:外展 π/2	矢状面:前屈 0
D	矢状面:前屈 2.845	矢状面:前屈 0
E	矢状面:前屈 π/2	矢状面:前屈 0
F	冠状面:外展 π/2 水平面:水平屈曲 π/4 矢状面:前屈至极限位置	矢状面:前屈 0
G	冠状面:外展 π/2 水平面:水平屈曲 π/4	矢状面:前屈 0

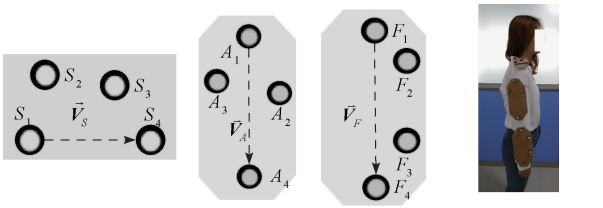


图 4　基于 VICON 的人体手臂达点运动实验

Fig. 4　Experimental figures of human arm reaching motion based on VICON

其次,把各刚体标志块按图 4(d)的方式贴于实验者的右臂。最后,为了保证所提取的特征的一般性和鲁棒性,对图 3 中的 14 个目标点分别做 100 次达点运动实验,并选用不同的实验者进行实验。其中,每次运动时手臂的初始位置为自然垂直向下,并采用 VICON 系统实时记录人体手臂达点运动时,各 marker 点和目标点的位置信息。

2.3　人体手臂关节变量构建

为了描述人体手臂各关节的运动情况,本文构建了肩夹角 α_1、肘夹角 α_2 和腕关节运动角 β_1,如图 5 所示。通过对这些人臂关节变量的分析与学习,能够提取出体现人体手臂运动的规律和约束。

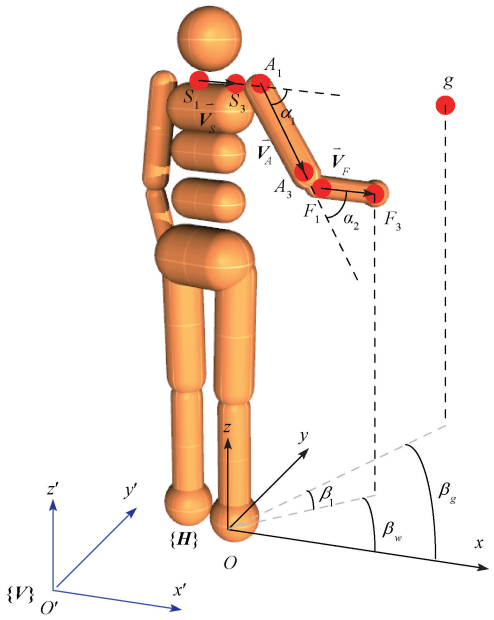


图 5　人体手臂关节变量

Fig. 5　Joint variables of human arm

肩夹角 α_1 由肩向量 $\vec{V}_S$ 和上臂向量 $\vec{V}_A$ 组成,它不仅体现了肩关节的运动情况,还描述了肩部和上臂的运动关系。

$$\begin{cases}\vec{V}_S = (x_{S4} - x_{S1}, y_{S4} - y_{S1}, z_{S4} - z_{S1}) \\ \vec{V}_A = (x_{A4} - x_{A1}, y_{A4} - y_{A1}, z_{A4} - z_{A1})\end{cases} \tag{1}$$

$$\alpha_1 = \arccos(\vec{V}_S \cdot \vec{V}_A / |\vec{V}_S| \cdot |\vec{V}_A|) \tag{2}$$

其中,x_i,y_i 和 z_i 分别为点 i 在人体坐标系 $\{H\}$ 下的位置坐标。

同样的,肘夹角 α_2 由上臂向量 $\vec{V}_A$ 和前臂向量 $\vec{V}_F$ 组成,它不仅体现了肘关节的运动情况,还能够描述上臂和前臂的运动关系。

$$\vec{V}_F = (x_{F4} - x_{F1}, y_{F4} - y_{F1}, z_{F4} - z_{F1}) \tag{3}$$

$$\alpha_2 = \arccos(\vec{V}_A \cdot \vec{V}_F / |\vec{V}_A| \cdot |\vec{V}_F|) \tag{4}$$

腕关节运动角 β_1 则描述了腕关节与目标点的运动情况。在人体坐标系 $\{H\}$ 下,设腕关节 $F4$ 的投影点与坐标原点相连形成的向量和 x 轴的夹角为 β_w,目标点 $\boldsymbol{g}=(x_g, y_g, z_g)$ 的投影点与坐标原点形成的向量和 x 轴的夹角为 β_g。腕关节运动角 β_1 即为 β_w 与 β_g 的角度差。

$$\beta_1 = \left| \arctan\left(\frac{y_g}{x_g}\right) - \arctan\left(\frac{y_{F4}}{x_{F4}}\right) \right| \tag{5}$$

2.4　数据分析与学习

由于肩夹角 α_1 和肘夹角 α_2 能够描述人臂运动过程中各部位间的相互关系,因此人臂各部位间的制约关系可以通过 α_1 和 α_2 的运动范围来体现。图 6 为以点 G 为目标点做 100 次达点运动时,肘夹角 α_2 的最小值的频率直方图。由于该频率直方图接近于概率密度曲线,因此能够初步判断这组数据满足正态分布。图 7 为对应的肘夹角 α_2 的最小值的 Q-Q (Quantile-Quantile)图。由于 Q-Q 图的判定依据为样本分位数与理论分位数形成的数据大致呈现为一条直线的状态,因此能够进一步确认该组数据具备正态分布的特点。

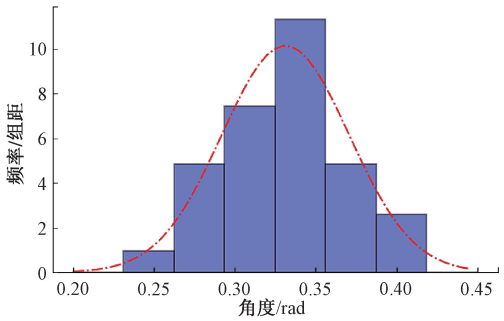


图 6　肘夹角 α_2 的最小值的频率直方图

Fig. 6　Frequency histogram of the minimum value of elbow angle α_2

最后,利用 Shapiro-Wilk 假设检验对该组数据进行检验计算。Shapiro-Wilk 检验是一种显著性假设检验方法,从统计学意义上将样本分布与正态分布进行比较,以

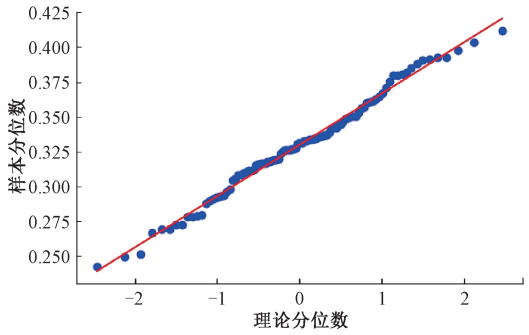


图 7 肘夹角 α_2 的最小值的的 Q-Q 图

Fig. 7 Q-Q diagram of minimum value of elbow angle α_2

确定数据是否显示出与正态性的偏离或符合。其中，α_2 的最小值的 p-value = 0.552。由于该 p-*value* 明显大于显著性水平 α = 0.05，最终确定该组数据与正态分布没有明显区别。因此，对于目标点 G，能够采用参数估计的方法获得肘夹角 α_2 的最小值。

同理，参考上述方法，首先对各目标点的 α_1 和 α_2 的最值进行正态性判断。结果表明，它们都具备正态分布的特性。然后，通过参数估计的方法求得各个目标点的 α_1 和 α_2 的最值，即 α_1 的范围 $\boldsymbol{M}_s$ 为[0.451，1.385] rad，α_2 的范围 $\boldsymbol{M}_e$ 为[0.138，1.906] rad。由于针对这些目标点的运动能够代表人体手臂在 $\boldsymbol{S}_M$ 运动空间中的运动情况，因此，在对 $\boldsymbol{S}_M$ 空间中任意目标点进行达点运动时，人体手臂肩夹角 $\alpha_1 \in \boldsymbol{M}_s$，肘夹角 $\alpha_2 \in \boldsymbol{M}_e$。

本文利用腕关节运动角 β_1 进一步研究了腕关节的运动状态。图 8 所示为目标点 G 某一次达点运动时 β_1 的变化情况，能够初步判断 β_1 和腕关节与目标点的距离呈正相关关系。在统计学中，通常使用 Spearman 相关系数度量非线性数据之间的相关性程度。其中，目标点 G 的 Spearman 相关系数为 0.996。该结果表明，对于目标点 G 而言，β_1 和腕关节与目标点的距离具备强正相关性。

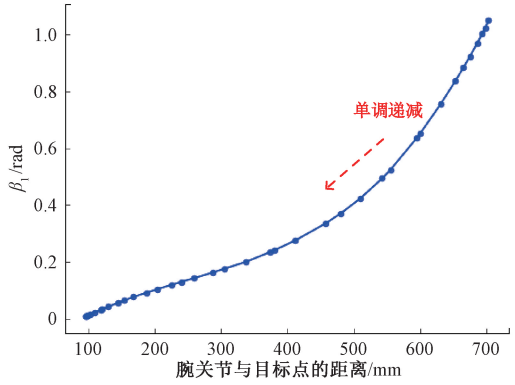


图 8 腕关节运动角 β_1 随距离的变化关系

Fig. 8 Relationship between wrist motion angle β_1 and distance

同理，对所有目标点进行相关性分析，能够发现，β_1 与距离存在强正相关关系。随着腕关节位置与目标点位置之间距离的减小，β_1 逐渐减小至某一范围。同时，该过程是单调递减的。根据表 2 中到达各目标点时 β_1 的收敛角度，能够确定到达目标点时 β_1 的收敛范围 $\boldsymbol{M}_w$ 为[0，0.013] rad，它体现了到达目标点位置时腕关节的可选运动状态。

表 2 腕关节运动角 β_1 的收敛情况

Table 2 Converging values of wrist joint motion angle β_1

rad

目标点	收敛角度	目标点	收敛角度
A	0.008	Random 1	0.013
B	0.006	Random 2	0.011
C	0.009	Random 3	0.010
D	0.011	Random 4	0.007
E	0.010	Random 5	0.011
F	0.009	Random 6	0.013
G	0.012	Random 7	0.012

3 强化学习 HP 算法

本文采用物理模拟器 MuJoCo 构建强化学习环境，包括 Jaco2 机械臂、目标点和桌面，机械臂的状态为每次训练的初始状态，环境中的目标点和机械臂各关节的位置信息均以{B}为参考坐标系。

基于 DDPG 和 HER 强化学习算法，本文设计了仿人规划 HP 算法。该算法首先结合第 1.1 节中机械臂各关节的仿人运动范围，初步避免了机械臂的非拟人运动规划。然后，结合任务目标和提取的人体手臂达点运动的动作特征，设计对应的仿人回报函数，进一步保证了运动规划的仿人性。强化学习过程中，回报函数不仅能够用于衡量任务实现的好坏，还能够简化任务的训练。若要实现机械臂仿人运动的训练，仿人回报函数的设计是 HP 算法的关键。设 r_t 为每回合的即时回报，本文将从目标点约束、仿人运动约束和仿人姿态约束 3 个方面设计仿人回报函数。

对机械臂仿人运动而言，首要任务是使机械臂的末端朝目标点运动，同时避免发生碰撞。若检测到机械臂发生碰撞，则直接结束本回合的训练，并将即时回报设置为 r_t = − 100。当机械臂不发生碰撞时，以机械臂末端与目标点的距离为参考，回报函数 r_1 如式(7) 所示。

$$dis = \| \boldsymbol{X}_E - \boldsymbol{X}_g \| \cdot K \tag{6}$$

$$r_1 = \begin{cases} -\ln(dis), & dis > 1 \\ 0, & dis \leqslant 1 \end{cases} \tag{7}$$

其中,dis 为机械臂末端与目标点的距离,$\boldsymbol{X}_E \in \boldsymbol{R}^3$ 为机械臂末端的位置,$\boldsymbol{X}_g \in \boldsymbol{R}^3$ 为目标点的位置,$\|\cdot\|$为欧几里得距离。由于强化学习仿真环境中获取的机械臂的位置信息以米为单位,为了避免因单位过大导致的较小的数值变化,本文将距离信息 dis 的阈值设定为以厘米为单位,即 $K = 100$。式(7)的第 1 部分为当机械臂末端与目标点的距离大于 1 cm 时,则认为机械臂没有到达目标点位置,不满足目标点约束,此时的回报函数 r_1 为负回报,r_1 的值随着机械臂末端与目标点距离的增大而减小。第 2 部分为当距离小于 1 cm 时,则认为机械臂到达目标点位置,满足目标点约束,对应的 r_1 为 0。

根据第 2.4 节的分析结果可知,人体手臂在对 $\boldsymbol{S}_M$ 运动空间中的任意一个目标点进行达点运动时,其对应的肩夹角和肘关节在运动过程中满足 $\alpha_1 \in \boldsymbol{M}_s$, $\alpha_2 \in \boldsymbol{M}_e$。针对仿人运动约束的回报函数,设 $\boldsymbol{X}_i \in \boldsymbol{R}^3, i \in [1,6]$ 分别对应强化学习中关节 1~6 的位置,由式(2)和(4)可实时计算出机械臂的肩夹角 α_{1r} 和肘夹角 α_{2r}。因此,对于强化学习每一时间步下的机械臂而言,当 $\alpha_{1r} \in \boldsymbol{M}_s$, $\alpha_{2r} \in \boldsymbol{M}_e$ 时,说明当前机械臂的运动满足肩夹角和肘夹角约束,具备仿人性,此时回报函数 $r_2 = 0$。当 $\alpha_{1r} \notin \boldsymbol{M}_s$, $\alpha_{2r} \notin \boldsymbol{M}_e$ 时,说明当前机械臂的运动不具备仿人性,其回报函数 r_2 如式(8)所示,随着肩夹角、肘夹角与 $\boldsymbol{M}_s$、$\boldsymbol{M}_e$ 偏离程度的加剧,r_2 的值越来越小。

$$r_2 = -\omega_1 \cdot \frac{\arctan\Delta\alpha_1}{\pi/2} - \omega_2 \cdot \frac{\arctan\Delta\alpha_2}{\pi/2} \tag{8}$$

其中,$\Delta\alpha_1$ 为机械臂当前肩夹角 α_{1r} 与人臂肩夹角范围的偏差,$\Delta\alpha_2$ 为机械臂当前肘夹角 α_{2r} 与人臂肘夹角范围的偏差,ω_1 和 ω_2 为表示重要性程度的权重。

对于 Jaco2 机械臂的腕关节运动而言,相较于关节 4 和 6,关节 5 对其影响较大,因此设机械臂关节 5 的腕关节运动角为 β_{1r}。根据第 2.4 节可知,仿人姿态约束不仅需要判断运动过程中 β_{1r} 的单调性,还需要判断 β_{1r} 的收敛情况。由于强化学习是一步步进行的,β_{1r} 的单调性可以通过比较每一时间步下的机械臂的腕关节运动角 β_{1r} 实现。若在第 t 时间步下,β_{1r} 不满足单调递减的特性,则说明机械臂存在非拟人运动,如机械臂的左右绕动,此时的回报函数 r_3 应为 -1。若在第 t 时间步下,β_{1r} 满足单调递减的特性,则需要进一步根据人体手臂腕关节运动角 β_1 的收敛情况,判断机械臂运动的仿人性。当 $\beta_{1r} \in \boldsymbol{M}_w$ 时,机械臂满足腕关节运动角约束,具备仿人性,此时的回报函数 r_3 应为 0。当 $\beta_{1r} \notin \boldsymbol{M}_w$ 时,机械臂不满足腕关节运动角约束,不具备仿人性,对应的仿人回报函数 r_3 如式(9)所示,随着腕关节运动角 β_{1r} 与收敛范围 $\boldsymbol{M}_w$ 偏离程度的加剧,r_3 的值越来越小。

$$r_3 = -\frac{\arctan\beta_{1r}}{\pi/2} \tag{9}$$

为了避免设计的各项仿人回报函数在数值上存在较大的差异,本文需要将仿人运动约束和仿人姿态约束对应的回报函数的值归一化到 [-1, 0] 内。由于 $y = \arctan(x), x > 0$ 的数值范围为 $(0, \pi/2)$,因此,式(8)和(9)中除以 $\pi/2$ 的目的是将式中每一项的数值范围归一化至 $(-1, 0)$。同时,为了保证 r_2 的回报值在 $[-1, 0]$ 内,ω_1 和 ω_2 之和应为 1,它们分别体现了肩夹角和肘夹角的重要程度。对于肩夹角和肘夹角约束而言,肩夹角和肘夹角互不影响,分别体现了人体手臂不同部位的关系,因此,它们应该具备相同的重要性,即权重 $\omega_1 = \omega_2 = 0.5$。

总之,若强化学习的每次迭代中都满足上述 3 个约束,则即时回报 $r_t = 10$;若 3 个约束中有任意一个不满足,则即时回报 r_t 如式(10)所示。权重 $\lambda_1 \sim \lambda_3$ 不仅保证了主要奖励的核心地位,还避免了各种辅助回报的影响。

$$r_t = \sum_{i=1}^{3} \lambda_i \cdot r_i \tag{10}$$

其中,权重 $\lambda_1 = \lambda_2 = 1$ 和 $\lambda_3 = 0.5$。

4　机械臂仿人运动平台验证及分析

4.1　HP 算法仿真训练

本文采用 OpenAI 的强化学习仿真环境 Gym 进行算法测试和参数训练。为了获取机械臂仿人运动规划 HP 算法的最优训练结果,表 3 所示为不同更新次数和时间步数下的模型训练的成功率和耗时。根据表 3 中的训练结果能够得出:1)不同的更新次数和时间步数,对训练结果存在较大的影响;2)在同一时间步数下,随着更新次数的增加,训练时长明显增加,模型训练的成功率先升高后降低;3)在同一更新次数下,随着时间步数的增加,训练时长逐渐增加,模型训练的成功率先升高后降低;4)训练的回合数决定着策略更新的次数,过少的回合数将产生较少的更新次数,从而导致较差的训练结果;过多的回合数则会造成智能体的过度训练,进而降低任务的成功率;5)任务的完成情况受时间步数的影响,过少的时间步数将导致回合内的任务无法完成,从而降低任务的成功率;过多的时间步数则会造成任务的过度拟合,继而产生较差的训练结果。

结合表 3,通过对比不同的训练回合数和时间步数下机械臂仿人运动的成功率,本文最终将更新次数设为 200,即对应的训练回合数为 10 000,每回合的时间步数设为 20。此时,模型训练的成功率最高,且训练耗时最短。图 9 为对应的机械臂强化学习仿人运动的训练结果。随着更新次数的增加,训练的成功率平均为 90.8%。

表 3 不同更新次数和时间步数下的模型训练的成功率和耗时

Table 3 Success rate and time consumption of model training under different epochs and timesteps

更新次数	时间步数									
	timesteps = 10		timesteps = 20		timesteps = 30		timesteps = 40		timesteps = 50	
	成功率/%	耗时/s	成功率/%	耗时/s	成功率/%	耗时/s	成功率/%	耗时/s	成功率/%	耗时/s
epochs = 100	77.96	2 567.3	85.95	2 578.0	85.27	2 676.7	86.73	2 727.2	87.08	3 054.4
epochs = 150	77.84	4 052.8	87.08	4 080.1	87.43	4 175.5	86.97	4 218.0	85.88	4 409.6
epochs = 200	83.64	5 831.6	90.80	5 963.8	90.78	5 980.1	88.62	6 167.5	85.76	6 387.8
epochs = 250	82.15	7 633.8	89.35	7 678.7	89.20	7 780.7	87.58	8 180.6	85.56	8 198.0
epochs = 300	81.36	9 229.2	87.99	9415.2	89.10	9 550.3	86.91	9 835.4	84.86	10 331.5
epochs = 350	79.60	11 039.2	87.58	11 573.8	88.09	11 719.2	86.16	11 953.3	83.29	12 482.2

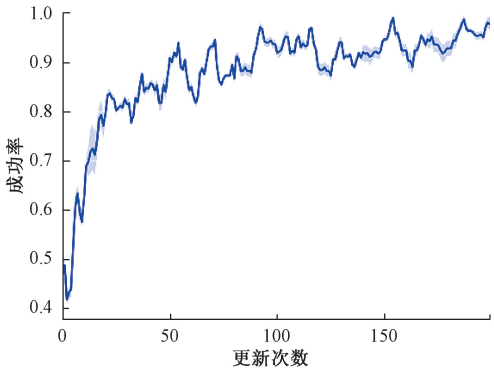


图 9 强化学习训练结果

Fig. 9 Training results of reinforcement learning

4.2 HP 算法的仿人性验证

机械臂仿人运动平台主要分为两个部分:仿人轨迹规划平台部分和机械臂运动控制平台部分。它们分别由 PC1 和 PC2 两台计算机控制,并通过 TCP/IP 通信协议共同实现真实机械臂的仿人运动规划。图 10 所示为机械臂仿人运动平台实物图。

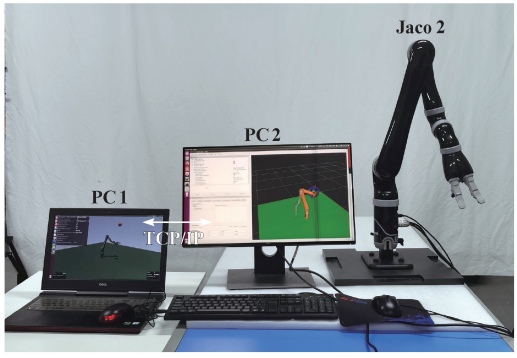


图 10 机械臂仿人运动平台

Fig. 10 The robotic arm humanoid motion platform

为验证 HP 算法的仿人性,本文基于机械臂仿人运动平台,针对机械臂的达点运动,设计了如下验证方案。首先,随机选取 4 个点作为机械臂达点运动的目标点。其中,4 个目标点分别位于机械臂基座坐标系 $\{B\}$ 的第Ⅰ~Ⅳ卦限中,用于代表机械臂在 $\{B\}$ 坐标系的上半空间中的达点运动情况。其次,分析达点运动过程中机械臂的运动轨迹和各关节的运动范围,初步检验规划算法的仿人性。最后,进一步分析由 HP 算法规划生成的机械臂运动是否满足第 2.4 节中对应的人体手臂运动规律,完成 HP 算法的仿人性检验。

随机选取 $\boldsymbol{g}_1 = (0.23, 0.42, 1.01)$, $\boldsymbol{g}_2 = (-0.4, 0.53, 0.63)$, $\boldsymbol{g}_3 = (-0.3, -0.57, 0.71)$ 和 $\boldsymbol{g}_4 = (0.3, -0.38, 0.85)$ 作为机械臂在 $\{B\}$ 坐标系的第Ⅰ~Ⅳ卦限中的目标点。表 4 所示为机械臂在上述达点运动过程中,各关节角的变化范围。通过对比第 1.1 节中机械臂各关节的仿人运动范围能够发现,在对上述目标点进行达点运动时,机械臂各关节的运动范围均满足仿人运动范围的要求,不存在超过人体手臂运动范围的运动。根据机械臂的上述各项运动指标,即运动轨迹和关节角变化范围,能够初步判定,通过 HP 算法规划的运动具备仿人性的特征。

为了进一步验证 HP 算法的仿人性,实验方案结合了第 2.4 节的内容,对达点运动过程中机械臂的肩夹角 α_{1r}、肘夹角 α_{2r} 和腕关节运动角 β_{1r} 进行了计算与分析。表 5 所示为机械臂在以 $\boldsymbol{g}_1 \sim \boldsymbol{g}_4$ 为目标点的达点运动过程中,肩夹角 α_{1r} 和肘夹角 α_{2r} 的变化范围。根据人体手臂肩夹角 α_1 的运动范围 $\boldsymbol{M}_s \in [0.451, 1.385]$ rad 和肘夹角 α_2 的运动范围 $\boldsymbol{M}_e \in [0.138, 1.906]$ rad,能够确定机械臂在上述达点运动过程中的肩夹角 $\alpha_{1r} \in \boldsymbol{M}_s$,肘夹角 $\alpha_{2r} \in \boldsymbol{M}_e$,满足人体手臂肩部、上臂与前臂的运动关系。

图 11 为机械臂在上述达点运动过程中腕关节运动角 β_{1r} 的变化情况。对于各个目标点而言,随着机械臂腕关节与目标点距离的减小,机械臂的腕关节运动角 β_{1r} 逐渐减小。同时,当机械臂到达目标点时,对应的 β_{1r} 均为 0,即 β_{1r} 的收敛值均在 $\boldsymbol{M}_w \in [0, 0.013]$ rad 内。结合第 2.4 节的分析结果,机械臂 β_{1r} 的变化情况不仅满足人

表 4　机械臂达点运动各关节角的变化范围

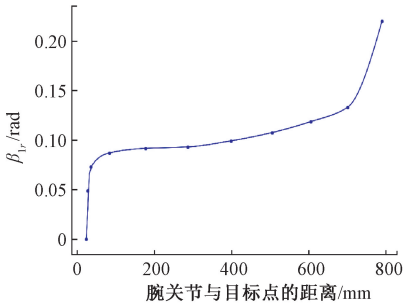
Table 4　Variation range of joint angles of robotic arm in reaching motion

目标点	运动范围/rad					
	关节 1	关节 2	关节 3	关节 4	关节 5	关节 6
$\boldsymbol{g}_1$	[1.571,2.140]	[3.438,3.491]	[0.559,2.289]	[0,0.652]	[−1.280,0]	[0,0.102]
$\boldsymbol{g}_2$	[0,1.571]	[3.438,3.601]	[0.559,1.576]	[0,0.375]	[−0.843,0]	[0,0.019]
$\boldsymbol{g}_3$	[−1.571,1.571]	[3.438,3.758]	[0.559,1.701]	[0,0.711]	[−1.005,0]	[0,0.094]
$\boldsymbol{g}_4$	[1.571,4.012]	[3.438,3.539]	[0.559,1.748]	[0,0.583]	[−1.186,0]	[0,0.037]

表 5　达点运动过程中机械臂 $\boldsymbol{\alpha}_{1r}$ 和 $\boldsymbol{\alpha}_{2r}$ 的变化范围

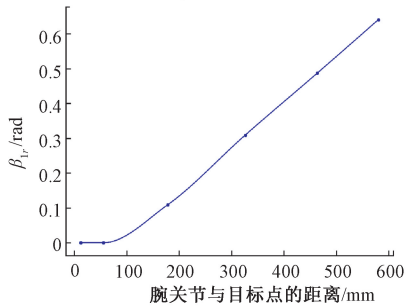
Table 5　Range of $\boldsymbol{\alpha}_{1r}$ and $\boldsymbol{\alpha}_{2r}$ of robotic arm during reaching motion

目标点	肩夹角 α_{1r} 变化范围/rad	肘夹角 α_{2r} 变化范围/rad
$\boldsymbol{g}_1$	[1.040,1.193]	[0.622,1.451]
$\boldsymbol{g}_2$	[0.980,1.093]	[0.730,1.567]
$\boldsymbol{g}_3$	[0.834,0.943]	[0.865,1.698]
$\boldsymbol{g}_4$	[1.091,1.232]	[0.812,1.563]



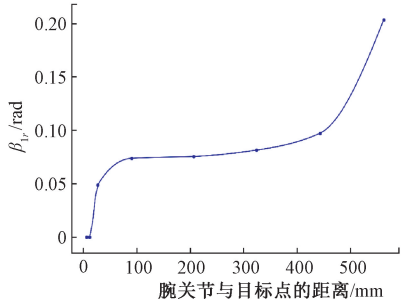
(a) 以g_1为目标点

(a) g_1 as the target point



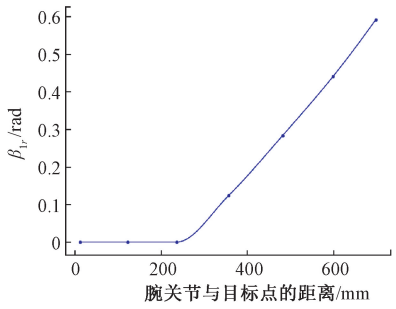
(b) 以g_2为目标点

(b) g_2 as the target point



(c) 以g_3为目标点

(c) g_3 as the target point



(d) 以g_4为目标点

(d) g_4 as the target point

图 11　机械臂达点运动过程中腕关节运动角的变化情况

Fig. 11　Wrist joint motion angle in the process of robotic arm reaching motion

体手臂达点运动过程中,腕关节运动角β_1随距离减小而单调递减的要求,还满足到达目标点时$\beta_1 \in \boldsymbol{M}_w$的要求。通过分析机械臂对上述目标点做达点运动时的动作特征,即肩夹角α_{1r}、肘夹角α_{2r}和腕关节运动角β_{1r},能够最终判定,HP 算法规划的运动具备仿人性。

因此,对于$\{B\}$坐标系上半空间中的任意目标点,通过 HP 算法规划的机械臂运动满足人体手臂的运动规律和约束条件,具备仿人性。

4.3　机械臂仿人运动结果分析

本文设计了如下实验,以分析 HP 算法仿人运动规划的成功率和强化学习中轨迹规划的耗时。首先,将实验按照$\{B\}$坐标系上半空间的 4 个卦限分为 4 个部分,每个部分对应一个卦限。在每个卦限中,随机选取 200 个点作为达点运动的目标点。其次,对每个目标点分别使用 HP 算法,以完成机械臂的达点运动规划,并记录对应达点运动的耗时。然后,统计并评估机械臂在每一卦限下的达点运动情况。最后,通过分析机械臂的达点运动结果,完成仿人规划成功率和耗时的计算。

强化学习中时间步的更新时长能够自行设置,较短的更新时长将缩短强化学习轨迹规划的耗时,因此不能直接将规划耗时作为判断轨迹规划耗时的标准。但由于时间步数不受更新时长的影响,故本文选择完成轨迹规划所需的时间步数作为强化学习轨迹规划的耗时指标。

表 6 所示为目标点在不同卦限下,机械臂仿人路径规划的成功率和耗时情况。

表 6 机械臂在不同卦限下仿人路径规划的成功率

Table 6 Success rate of humanoid planning for robotic arm in different octant

目标点位置	达点次数	成功次数	成功率/%	轨迹规划所需时间步数	机械臂达点运动耗时/s
第 I 卦限	200	185	92.5	7.46	4.67
第 II 卦限	200	180	90.0	7.62	9.37
第 III 卦限	200	183	91.5	7.51	15.19
第 IV 卦限	200	182	91.0	7.39	9.21
平均值			91.25	7.50	

结果表明,机械臂在每个卦限下的达点运动都具备仿人性,且每一卦限下的成功率和仿人轨迹规划耗时的差距并不大。从整体数据来看,机械臂仿人运动规划的成功率为 91.25%,规划轨迹所需的时间步数平均为 7.50 步。若不考虑机械臂关节 1 旋转运动带来的时间损耗,机械臂的仿人达点运动耗时合理。因此,HP 算法具备较高的仿人达点运动规划的成功率和较短的仿人规划耗时,验证了 HP 算法的可行性和有效性。

5 结 论

本文主要通过设计描述人体手臂各部位间相互作用的人臂关节变量,即肩夹角 α_1、肘夹角 α_2 和腕关节运动角 β_1,提出了一种基于强化学习的机器人手臂仿人运动规划方法。采用 VICON 系统对人体手臂三维空间的达点运动进行数据采集,并计算对应人臂关节变量。通过正态性和相关性分析的方式,获得手臂运动过程中的特征及约束条件,进而针对仿人约束的不同特点,设计相应的回报函数,保证强化学习的有效训练。通过大量实验,对比机械臂仿人运动和人臂运动的特征,本文所提方法生成的动作路径不仅在运动方式上满足人体运动的特性,在运动姿态上也具备仿人性。

参考文献

[1] FICUCIELLO F, MIGLIOZZI A, LAUDANTE G, et al. Vision-based grasp learning of an anthropomorphic hand-arm system in a synergy-based control framework [J]. Science Robotics, 2019, 4(26):1-11.

[2] KULIC D, VENTURE G, YAMANE K, et al. Anthropomorphic movement analysis and synthesis: A survey of methods and applications [J]. IEEE Transactions on Robotics, 2016, 32(4):776-795.

[3] 赵京, 郭兴伟, 谢碧云. 达点臂姿预测指标与机械臂仿人运动[J]. 机械工程学报, 2015, 51(23):21-27.
ZHAO J, GUO X W, XIE B Y. Criterion for human arm in reaching tasks and human-like motion planning of robotic arm [J]. Journal of Mechanical Engineering, 2015, 51(23):21-27.

[4] 陈盛,郃春,徐国政,等. 基于分数阶阻抗控制的 7 自由度机器人辅助主动康复训练方法[J]. 仪器仪表学报,2020, 41(9):196-205.
CHEN SH, TAI CH, XU G ZH, et al. 7-DOF robot-assisted active rehabilitation training method based on fractional impedance control [J]. Chinese Journal of Scientific Instrument, 2020, 41(9):196-205.

[5] LI M, GUO W, LIN R F, et al. An efficient motion generation method for redundant humanoid robot arm based on the intrinsic principles of human arm motion[J]. International Journal of Humanoid Robotics, 2018, 15(6): 1-20.

[6] GONG S Q, ZHAO J, ZHANG Z Q, et al. Task motion planning for anthropomorphic arms based on human arm movement primitives [J]. Industrial Robot-the International Journal of Robotic Research and Application, 2020, 47(5):669-681.

[7] GARCÍA N, ROSELL J, SUAREZ R. Motion planning by demonstration with human-likeness evaluation for dual-arm robots [J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2019, 49(11):2298-2307.

[8] ROSELL J, SUAREZ R, GARCIA N, et al. Planning grasping motions for humanoid robots [J]. International Journal of Humanoid Robotics, 2019, 16(6):1950041.

[9] YANG C, ZENG C, FANG C, et al. A DMPs-based framework for robot learning and generalization of humanlike variable impedance skills [J]. IEEE/ASME Transactions on Mechatronics, 2018, 23(3):1193-1203.

[10] ZHU Y, YANG C, WEI Q, et al. Human-robot shared control for humanoid manipulator trajectory planning[J]. Industrial Robot-The International Journal of Robotics

Research and Application, 2020, 47(3):395-407.

[11] WEI Y, JIANG W, RAHMANI A, et al. Motion planning for a humanoid mobile manipulator system[J]. International Journal of Humanoid Robotics, 2019, 16(2):1950006.

[12] DUARTE N F, RAKOVIC M, SANTOS-VICTOR J. Biologically inspired controller of human action behaviour for a humanoid robot in a dyadic scenario[C]. IEEE EUROCON 2019-18th International Conference on Smart Technologies, 2019:1-6.

[13] HU H M, DU J M, HU X H, et al. The experimental research on joint range of motion[M]. Advances in Physical Ergonomics and Human Factors, 2016, 489: 265-274.

[14] GATES D H, WALTERS L S, COWELEY J, et al. Range of motion requirements for upper-limb activities of daily living[J]. The American Journal of Occupational Therapy, 2015, 70(1):1-10.

[15] QIU C, HU Y, CHEN Y, et al. Deep deterministic policy gradient (DDPG)-based energy harvesting wireless communications[J]. IEEE Internet of Things Journal, 2019, 6(5):8577-8588.

[16] PRIANTO E, KIM M S, PARK J H, et al. Path planning for multi-arm manipulators using deep reinforcement learning: soft actor-critic with hindsight experience replay[J]. Sensors, 2020, 20(20): 5911-5932.

作者简介

杨傲雷(通信作者),2004 年于湖北工业大学获得学士学位,2009 年于上海大学获得硕士学位,2012 年于英国女王大学获得博士学位,现为上海大学副教授,主要研究方向为智能机器人与视觉学习系统、人体姿态与动作识别、多智能体协同控制等。

E-mail: aolei@shu.edu.cn

Yang Aolei received his B. Sc. degree from Hubei University of Technology in 2004, received his M. Sc. degree from Shanghai University in 2009, and received his Ph. D. degree from Queen's University Belfast in 2012. He is currently an associate professor at Shanghai University. His main research interests include intelligent robot and vision learning system, human posture and action recognition, and multi-agent cooperative control, etc.

徐昱琳,1986 年于东华大学获取学士学位,2003 年于法国斯特拉斯堡大学获得博士学位,现为上海大学副教授,主要研究方向多变量工业系统的建模与控制、仿人灵巧手结构设计、建模与控制、服务机器人智能控制等。

E-mail: xuyulin@shu.edu.cn

Xu Yulin received her B. Sc. degree from Donghua University in 1986 and received her Ph. D. degree from University of Strasbourg in 2003. She is currently an associate professor at Shanghai University. Her main research interests include modeling and control of multi-variable industrial systems, bionic manipulator structure design, modeling and control and service robot intelligent control.