Journal of Computer Applications
计算机应用,2013,33(5):1301 - 1304
ISSN 1001-9081
CODEN JYIIDU
2013-05-01
http://www.joca.cn

文章编号:1001 - 9081(2013)05 - 1301 - 04
doi:10.3724/SP.J.1087.2013.01301

基于分层 Option 的仿人机器人相似性关键姿势转换

柯文德[1,2*],彭志平[1],陈　珂[1],项顺伯[1]
(1.广东石油化工学院 计算机与电子信息学院,广东 茂名 525000;　2.哈尔滨工业大学 计算机科学与技术学院,哈尔滨 150001)
(*通信作者电子邮箱 wendeke@163.com)

摘　要:针对运动捕获系统获取的人体运动轨迹固定、难以实现仿人机器人关键姿势转换问题,提出了一种基于分层 Option 学习的仿人机器人关键姿势相似性转换方法。构建多级关键姿势树状结构,从关节相似差异、时刻整体相似差异、周期整体相似差异等角度描述了关键姿势差异,引入分层强化 Option 学习方法,建立关键姿势与 Option 行为集,由关键姿势差异的累计奖励将 SMDP-Q 方法逼近最优 Option 值函数,实现了关键姿势的转换。实验验证了方法的有效性。

关键词:仿人机器人;分层强化学习;相似性;姿势
中图分类号: TP242.6　　**文献标志码**:A

Similar key posture transformation based on hierarchical Option for humanoid robot

KE Wende[1,2*], PENG Zhiping[1], CHEN Ke[1], XIANG Shunbo[1]
(1. *College of Computer and Electronic Information, Guangdong University of Petrochemical Technology, Maoming Guangdong*525000, *China*;
2. *School of Computer Science and Technology, Harbin Institute of Technology, Harbin Heilongjiang* 150001, *China*)

Abstract: Concerning the problem in which the fixed locomotion track captured from human movement can not be used in transformation between key postures for humanoid robot, a method of similar key posture transformation based on hierarchical Option for humanoid robot was proposed. The multi-level dendrogram of key postures was constructed and the difference of key postures was illustrated in respects of similar joint difference, moment total similar difference, period total similar difference. The hierarchical reinforcement Option learning was introduced, in which the sets of key postures and Option actions were constructed. SMDP-Q method tended to be the optimal Option function by the accumulative rewards of key posture difference and the transformations were realized. The experiments show the validity of the method.

Key words: humanoid robot; hierarchical reinforcement learning; similarity; posture

0　引言

仿人机器人具有与人相似的躯干关节,能较好地模仿人体运动,是机器人研究领域的重要组成部分。仿人机器人关节运动轨迹的传统设计方法是通过求解运动解析方程实现的,具有运动轨迹平滑的优点,但与人类相比,在运动轨迹的自然过渡、能量消耗、复杂动作设计等方面存在很大差距[1]。基于人体运动相似性的仿人机器人动作设计是通过将捕获到的人体关节运动数据经由逆运动学解算后,施加运动学与动力学约束并应用于仿人机器人动作设计上,是目前仿人机器人的研究热点[2-3]。

仿人机器人相似性运动主要通过关键姿势转换实现,目前对关键姿势的研究主要体现在轨迹获取与跟踪方面,例如,文献[4]分析了由单摄像机捕获的网球击打运动过程并提出了关键姿势的自动识别与划分机制;文献[5]对人体姿势建立分类模型空间,在场景模型驱动下将连续运动轨迹划分为的多个单个分段,分别提取出关键姿势;文献[6]由几何法分析人体运动轨迹,构建出基于可变形三角形的关键姿势提取方法;文献[7]将二维人体运动捕获数据转换为匹配的三维关键运动,构建出空间运动轨迹并由行为分类器识别出关键姿势,等等。

以上研究中,关键姿势之间的变换依据连续、固定的轨迹进行,当在已有的关键轨迹中插入新的关键姿势时,运动轨迹的一致性受到破坏。为解决该问题,本文提出一种基于分层 Option 学习的关键姿势轨迹变换方法,采用 Option 分层强化学习中自适应、分层最优特点,将关键姿势定义为子状态,通过 Option 操作,反复迭代关节运动参数,实现从已有关键姿势插入新关键姿势的变换。

1　相似性运动特征描述

1.1　关键姿势划分

由于人体运动具有周期性特点,将该周期内的关节轨迹变化划分为若干个关键姿势,相邻关键姿势形成运动子相,从而形成相似性运动片段,对片段进行组合即形成不同的运动形态。关键姿势的划分原则为:若演示人体在某时刻相应关节停顿,则对应关节形成关键姿势。以左脚为例,构建出如图 1 所示的自下而上的关键姿势树状结构,其中,$\dot{\varphi}_i^R(t)$ 为 t 时刻机器人第 i 关节角速度,L 表示 left、R 表示 right,Hi 表示

收稿日期:2012-12-04;**修回日期**:2013-01-04。
基金项目:国家自然科学基金资助项目(61272382);广东省自然科学基金资助项目(8152500002000003, S2012010009963);广东省高等学校科技创新项目(2012KJCX0077);广东高校石化装备故障诊断与信息化控制工程中心项目(512009)。
作者简介:柯文德(1976 -),男,广东茂名人,博士研究生,副教授,CCF 会员,主要研究方向:机器人、计算机系统结构;　彭志平(1969 -),男,福建泉州人,教授,博士,主要研究方向:智能主体、机器人;　陈珂(1964 -),男,黑龙江牡丹江人,副教授,硕士,主要研究方向:多机器人协作、数据挖掘;　项顺伯(1979 -),男,安徽枞阳人,讲师,硕士,主要研究方向:计算机软件、机器人。

hip,Kn 表示 knee,An 表示 ankle,Fir 表示 first toe,Sec 表示 second toe,Thi 表示 third toe,For 表示 forth finger,Las 表示 last toe,则一级关键姿势周期为 $T_1 = t_{1\_2} - t_{1\_1}$,且

$$T_1 = \bigcup_{\gamma=1}^{L}(t_{\gamma\_2} - t_{\gamma\_1}) = \bigcup_{\gamma=1}^{L}\bigcup_{\eta=1}^{M_\gamma}(t_{\gamma\_\eta\_2} - t_{\gamma\_\eta\_1}) \tag{1}$$

其中:L 为最大层数,M_γ 为第 γ 层中最大节点数,t_{γ_2}、t_{γ_1} 分别为第 γ 层所有关键姿势中的最早起始时刻与最后结束时刻,$t_{\gamma\_\eta\_2}$、$t_{\gamma\_\eta\_1}$ 分别为第 γ 层第 η 个节点的最早起始时刻与最后结束时刻,则任一级关键姿势周期 $T_\gamma = t_{\gamma\_2} - t_{\gamma\_1}$,且

$$T_\gamma = \bigcup_{\eta=1}^{M_\gamma}(t_{\gamma\_\eta\_2} - t_{\gamma\_\eta\_1}) \tag{2}$$

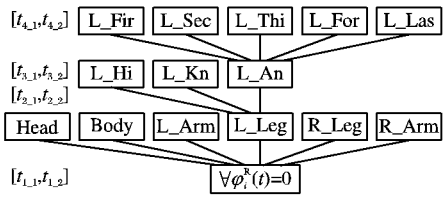

图 1 关键姿势树状图

在树状结构中,同级节点互为兄弟,具有相同优先级别,规定多级关键姿势优先级别从高到低满足以下规则:1)臂部规则,finger →wrist→elbow→shoulder;2)腿部规则,toe→ankle→knee→hip;3)躯干运动,neck→waist。可见,关键姿势的级别越高,姿势分解效果越明显。当在关键姿势间插入同级的新关键姿势时,运动周期缩短,相邻运动子相相似程度趋高。

1.2 相似性描述

考虑到演示人体与自然人之间的对应肢体比例差异,采用关节角度进行相似性描述,以较好地复现出人体运动特点,设定演示人体运动特征为标准数据,机器人运动特征为目标数据。设演示人体关节角度、角速度、角加速度分别为 $\varphi^H(t)$、$\dot{\varphi}^H(t)$、$\ddot{\varphi}^H(t)$,机器人对应关节角度、角速度、角加速度分别为 $\varphi^R(t)$、$\dot{\varphi}^R(t)$、$\ddot{\varphi}^R(t)$,关节角数量为 N。对于关键姿势差异,有如下定义:

定义 1 关节相似差异。设运动周期 T 内演示人体第 i 关节转角运动特征矢量和为 $\boldsymbol{C}_i^H = \sum_{j=1}^{M}\boldsymbol{C}_i^H(j)$,机器人对应关节特征矢量和为 $\boldsymbol{W}_i^R = \sum_{j=1}^{M}\boldsymbol{W}_i^R(j)$,其中 M 为 T 内相似性关键姿势数量,则运动周期 T 内第 i 关节的相似差异值为 $\Delta\boldsymbol{C}_i = \boldsymbol{C}_i^H - \boldsymbol{W}_i^R$。

定义 2 时刻整体相似差异。设 t 时刻演示人体整体关节转角运动特征矢量和为 $\boldsymbol{C}^H(t) = \boldsymbol{L}_1(\varphi_1^H(t),\dot{\varphi}_1^H(t),\ddot{\varphi}_1^H(t)) + \cdots + \boldsymbol{L}_N(\varphi_N^H(t),\dot{\varphi}_N^H(t),\ddot{\varphi}_N^H(t))$,机器人对应运动特征矢量和为 $\boldsymbol{W}^R(t) = \boldsymbol{W}_1^R(t) + \boldsymbol{W}_2^R(t) + \cdots + \boldsymbol{W}_N^R(t)$,则 t 时刻整体相似差异值为 $\Delta\boldsymbol{C}(t) = \boldsymbol{C}^H(t) - \boldsymbol{W}^R(t)$。

定义 3 周期整体相似差异。设运动周期 T 内演示人体整体关节转角运动特征矢量和为 $\boldsymbol{C}^H = \sum_{j=1}^{M}\boldsymbol{C}^H(j)$,机器人对应运动特征矢量和为 $\boldsymbol{W}^R = \sum_{j=1}^{M}\boldsymbol{W}^R(j)$,则整体相似性差异值为 $\Delta\boldsymbol{C} = \boldsymbol{C}^H - \boldsymbol{W}^R$。

由以上定义,若相似差异值趋向 0 时,描述演示人体与机器人具有趋同的运动特征,即满足式(3)最小时,关键姿势相似程度最大

$$S = \int_0^T\left(\sum_{i=1}^{N}\Delta\boldsymbol{C}_i\right)\mathrm{d}t \tag{3}$$

式(3)描述了在周期 T 内机器人实现人体相似运动过程的累计差异值,可见,将 S^{-1} 作为分层 Option 方法的累积奖励,促使从当前关键姿势状态转换到下一关键姿势的累积奖励值最大,即促使相似程度最大。

2 相似性关键姿势转换

2.1 分层 Option 方法

相对于传统强化学习策略具有的收敛速度慢、维数灾难等缺点,分层强化学习通过将问题空间分层与降维,在子空间内实现策略学习,提高了策略搜索速率。Swtton 对马尔可夫单步模型进行多时间步泛化,实现了 Option 下的强化学习[8]。Options 方法从学习任务中抽象出来并用于执行某子任务,描述为在某受限状态子空间中满足某策略约束的若干动作序列[9],动作可以是简单动作或者另一 Option,从而可构造出 Options 的分层关系,使得上层 Option 可调用下层 Option,具体 Option 的产生可以来源于先验知识或者自动生成[10-11]。

在分层强化学习(Hierarchical Reinforcement Learning,HRL)中定义 Markov-Option 与 Semi-Markov-Option[12]。

Markov-Option:在马尔可夫决策过程(Markov Decision Process,MDP)定义 Option 时,可由 3 元组 $\langle I,\pi,T\rangle$ 描述。其中:$I \subseteq S$ 为 Option 入口状态集,包括 Option 所有可用状态;π:$S\times A\rightarrow[0,1]$ 为内部策略;T:$S\rightarrow[0,1]$ 为终止条件。启动 Option 后,由策略 π 选取动作执行并由 s' 依概率 $T(s)$ 进入终止状态,当 $T(s_G) = 1$ 时,表明满足子目标 s_G 终止条件。特别地,可将基本动作(Primary action)$a \in A$ 视为单步 Option。

Semi-Markov-Option:在半马尔可夫决策过程(Semi-Markov Decision Process,MDP)上启动 Option 后的历史状态、动作与奖赏等决定了内部策略 π 与终止条件 T,Option 定义同样可用 3 元组 $\langle I,\mu,T\rangle$ 来描述。其中:$I \subseteq S$ 为 Option 入口状态集,包括 Option 所有可用状态;内部策略为 μ:$\Omega\times A\rightarrow[0,1]$;终止条件为 T:$\Omega\rightarrow[0,1]$,Ω 为历史集。在激活初始 Option $\langle I,\mu,T\rangle$ 后,遵循策略 μ 执行 Option,并允许 Option 选择执行关联的 Option,重复该过程直到结束于终止条件 T。当将 Semi-Markov-Option 的 Option 集合叠加到 MDP,则生成离散时间的 SMDP。若在此过程反复展开 Option 直到基本动作层,将可依策略 μ 生成若干常规策略,该策略被 Sutton 定义为平坦策略 $flat(\mu)$。

2.2 基于 Option 方法的关键姿势转换

设机器人运动关键姿势状态集 $S = \{s_1,s_2,\cdots,s_i,\cdots,s_k\}$,Option 集 $O = \{o_1,o_2,\cdots,o_i,\cdots,o_u\}$,其中,$u$ 为关键姿势数量,$o_i = f(\varphi_i^R,\dot{\varphi}_i^R,\ddot{\varphi}_i^R)$ 描述运动 f 函数根据 i 关节运动参数调整运动方向与速度,用 $\varepsilon(o_i,s_i,t)$ 表示在时刻 t、关键姿势 s_i 下激活 Option o_i,则关键姿势转换奖赏函数 $r(s_i,o_i)$[12] 为:

$$r(s_{i+1},o_i) = E\{r_{t+1} + \gamma r_{t+2} + \cdots + \gamma^{N-1}r_{t+N} \mid \varepsilon(o_i,s_i,t)\} \tag{4}$$

关键姿势状态转换概率函数 $p(s_{i+1} \mid s_i,o_i)$ 为:

$$p(s_{i+1} \mid s_i,o_i) = \sum_{N=1}^{\infty}p(s_{i+1},\sigma)\gamma^\sigma \tag{5}$$

其中:r 为关键姿势转换的累计奖励,γ 为折扣值,$p(s_{i+1},\sigma)$ 描述在持续时间步 σ 内从关键姿势状态 s_i 到达 s_{i+1} 的概率。

将式(4)、(5)引入分层 Options 中,对应最优值函数与最

优值 Option 值函数的 Bellman 方程分别为:

$$V_O^*(s_i) = \max_o\left[r(s_i,o_i) + \sum_{s'} p(s_{i+1}\mid s_i,o_i)V_O^*(s_{i+1})\right] \tag{6}$$

$$Q_O^*(s_i,o_i) = r(s_i,o_i) + \sum_{s'} p(s_{i+1}\mid s_i,o_i) \max_{o'} Q_O^*(s_{i+1},o_{i+1}) \tag{7}$$

$r(s_i,o_i)$ 描述关键姿势转换 Option o_i 的累积奖励,引入式(3),则

$$r(s_i,o_i) = S^{-1} = \left[\int_0^{\tau}\left(\sum_{i=1}^{n}\Delta C_i\right)\mathrm{d}t\right]^{-1} \tag{8}$$

从而,第 $k+1$ 次迭代的 Q 学习函数为

$$Q_{k+1}(s_i,a) = (1-\alpha)Q_k(s_i,a) + \alpha[r_t + \gamma r_t + \cdots + \gamma^{t-1}r_{t+N-1} \max_{a'} Q_k(s',a')] \tag{9}$$

扩展式(8),可将 SMDP-Q 方法逼近最优 Option 值函数,即

$$Q_{k+1}(s_i,o_i) = (1-\alpha)Q_k(s_i,o_i) + \alpha_k\left[r + \gamma^N \max_{o'} Q_k(s_{i+1},o_{i+1})\right] \tag{10}$$

算法流程为:

1) 初始化 Q 值表;

2) 观察当前关键姿势状态 s 并执行动作 a;

3) 得到新状态 s' 与即时奖励 r;

4) 由式(7) 更新 Q 值并记录下学习过程(s,a,s',r);

5) 若在持续时间 T 内未能到达下一关键姿势则转 6),否则转 2);

6) 为每个关节生成 Option 入口状态集并进行内部策略学习;

7) 进行 SMDP 学习;

8) 观察当前姿势状态并选择关节 Option;

9) 执行 Option 内部策略动作,观察关节角度变化情况,并更新 Q 值表;

10) 若未能到达目标关键姿势转 7),否则在 Option 内部重新学习;

11) 若到达目标状态则停止学习,否则转 8)。

3　实验验证

机器人的关节一般为几十个,而人体运动关节有数百个,因此需通过运动重定向与模型简化将演示人体的各个关节映射到仿人机器人运动模型上。图 2、3 为演示人体将运动姿势重定向到仿人机器人 Aldebaran Nao 对应肢体关节。

如图 4 所示,按照多级关键姿势树状结构将演示人体举右手过程划分为同级的 8 个关键姿势状态,其中第一个与最后一个关键姿势被视为起始状态与终止状态,其余每个关键姿势均被视为中间插入状态,Option 学习任务是找到从起始状态经过中间插入状态到达最终状态的各关节最优运动轨迹策略。基本动作为上下左右,算法流程步骤 5) 中的 T 描述策略探索深度,由经验值设置为 3,当在 3 个周期内未确定新状态则计算生成 Option,生成关节角度策略探测概率为 0.3,折扣因子 $\gamma=0.8$,当手臂各关节到达目标关键位姿将获得奖励,奖励值为 100,否则获得惩罚值 -25。

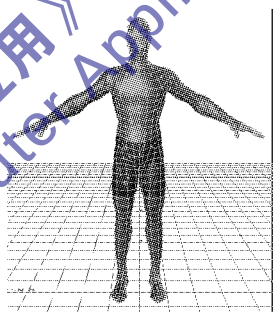

图 2　演示人体三维仿真效果

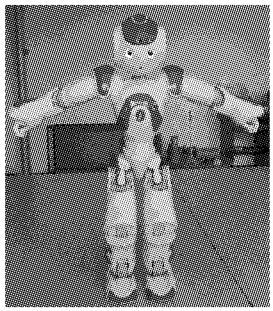

图 3　仿人机器人相似性运动

图 4　演示人体举右手

运行 Option 学习算法并依移动轨迹关节位姿匹配中间状态,当成功到达终止状态后重新回到初始状态学习。Option 学习算法运行 40 次,使用 ε 贪婪策略并设置学习率为 0.06,即 $\varepsilon=0.06$,并将相邻关键姿势转换时间设为时间步。图 5 为 Option 学习中状态平均 Q 值随时间步的变化情况,可见,在开始时间内学习效果较为缓慢,之后性能突然提高并收敛到最优解。图 6、图 7 所示为分别为根据 Option 学习策略实现的仿人机器人相似性举右手仿真与实际效果。

4　结语

基于人体运动的仿人机器人动作设计是目前机器人领域的研究热点,本文构造关键姿势树状结构并定义相似性变换的差异值,通过分层 Option 算法搜索从初始状态经中间插入状态并最终到达终止状态的关键姿势变换策略,实现了仿人机器人关键姿势转换的自调整自适应的寻优过程,并很好地保留了演示人体运动轨迹特点。

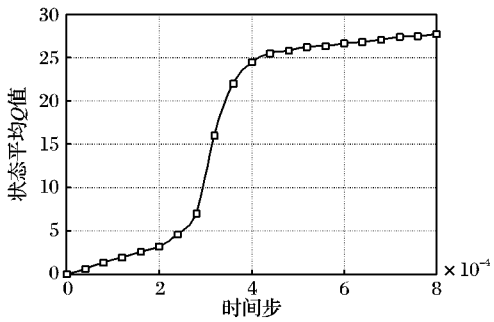

图 5　状态平均 Q 值变化

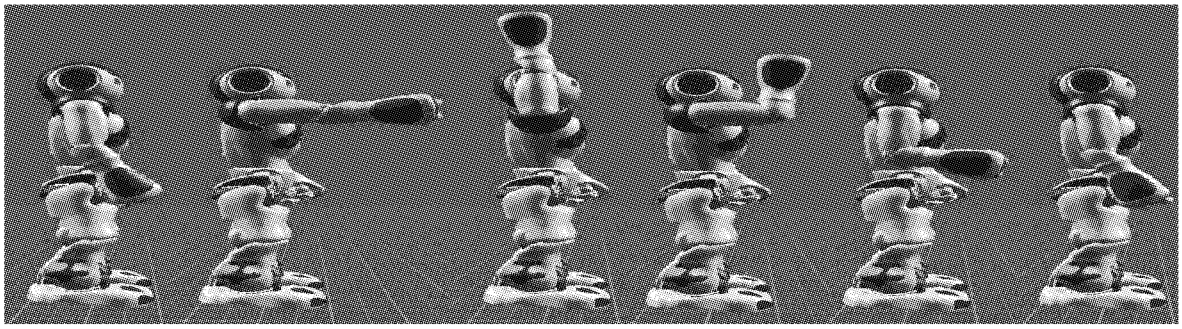

图 6 机器人举右手仿真图

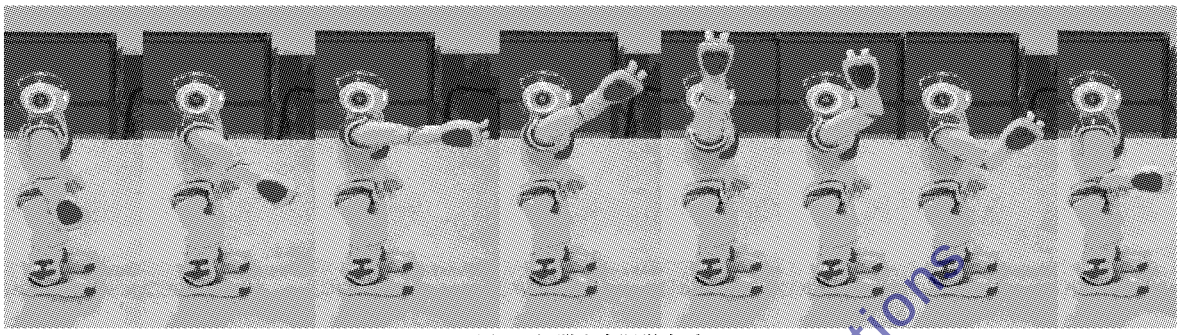

图 7 机器人实际举右手

参考文献：

[1] 赵晓军，黄强，彭朝琴，等. 基于人体运动的仿人型机器人动作的运动学匹配[J]. 机器人，2005，27(4)：358－361

[2] 柯文德，崔刚，洪炳镕，等. 参数化优化的仿人机器人相似性前向倒地研究[J]. 自动化学报，2011，37(8)：1006－1013.

[3] ARISTIDOU A，LASENBY J. Motion capture with constrained inverse kinematics for real-time hand tracking[C]// The 4th International Symposium on Communications，Control and Signal Processing. Piscataway：IEEE Press，2010：1－5.

[4] CONNAGHAN D，CONAIRE O，KEUY P，*et al*. Recognition of tennis strokes using key postures[C]// ISSC 2010：Signals and Systems Conference. Piscataway：IEEE，2010：245－248.

[5] HSIEH J W，CHEN S Y，CHUANG C H，*et al*. Occluded human body segmentation and its application to behavior analysis[C]// Proceedings of 2010 IEEE International Symposium on Circuits and Systems. Piscataway：IEEE，2010：3433－3436.

[6] HSIEH J W，CHUANG C H，CHEN S Y，*et al*. Segmentation of human body parts using deformable triangulation[J]. IEEE Transactions on Systems，Man and Cybernetics，Part A：Systems and Humans，2010，40(3)：596－610.

[7] 谷军霞，丁晓青，王生进. 基于人体行为3D模型的2D行为识别[J]. 自动化学报，2010，36(1)：46－53.

[8] SUTTON R S，PRECUP D，SINGH S. Between MDPs and semi-MDPs a framework for temporal abstraction in reinforcement learning[J]. Artificial Intelligence，1999，112(1)：181－211.

[9] DIETTERICH T G. Hierarchical reinforcement learning with the MAXQ value function decomposition[J]. Journal of Artificial Intelligence Research，2000，13(1)：227－303.

[10] BARTO A G，MAHADEVAN S. Recent advances in hierarchical reinforcement learning[J]. Discrete Event Dynamic Systems，2003，13(1/2)：41－77.

[11] 沈晶，刘海波，张汝波，等. 基于半马尔可夫对策的多机器人分层强化学习[J]. 山东大学学报：工学版，2010，40(4)：1－7.

[12] 沈晶. 分层强化学习方法研究[D]. 哈尔滨：哈尔滨工程大学，2006.

（上接第 1300 页）

[2] YUAN J，TANG G Y. Formation control for mobile multiple robots based on hierarchical virtual structures [C]// 2010 8th IEEE International Conference on Control and Automation. Piscataway：IEEE，2010：393－398.

[3] GHOMMAM J，MEHRJERDI H，SAAD M. Leader-follower formation control of nonholonomic robots with fuzzy logic based approach for obstacle avoidance [C]// 2011 IEEE International Conference on Intelligent Robots and Systems. Piscataway：IEEE，2011：2340－2345.

[4] 张玉礼，吴怀宇，程磊. 基于领航者模式的多机器人编队实现[J]. 信息技术，2010(11)：17－23.

[5] VIGURIA A，HOWARD A M. An integrated approach for achieving multirobot task formations [J]. IEEE Transactions on Mechatronics，2009，14(2)：176－186.

[6] 张捍东，黄鹂，岑豫皖. 改进的多移动机器人混合编队方法[J]. 计算机应用，2012，32(7)：1955－1964.

[7] NI J J，YANG S X. Bioinspired neural network for real-time cooperative hunting by multirobots in unknown environments [J]. IEEE Transactions on Neural Networks，2011，22(12)：2062－2077.

[8] YANG S X，MENG M. Neural network approaches to dynamic collision-free trajectory generation [J]. IEEE Transactions on Cybernetics，2001，31(3)：302－318.

[9] 张颖，陈雪波. 广义蚁群算法及其在机器人队形变换中的应用[J]. 模式识别与人工智能，2007，20(3)：319－324.

[10] 梁家海. 移动机器人编队的运动控制策略[J]. 计算机应用，2011，31(12)：3312－3314.

[11] QU H，YANG S X，WILLMS A R，*et al*. Real-time robot path planning based on a modified pulse-coupled neural network model [J]. IEEE Transactions on Neural Networks，2009，20(11)：1724－1739.

[12] 范莉丽，王奇志. 改进的生物激励神经网络的机器人路径规划[J]. 计算机技术与发展，2006，16(4)：19－21.