

平成 30 年 6 月 20 日現在

機関番号：55201

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K00355

研究課題名(和文) マルチタスク環境におけるCPGと強化学習を用いた多脚ロボットの行動獲得

研究課題名(英文) Behavior Acquisition of Multi-legged Robots by CPG and Reinforcement Learning in Multi-task Environment

研究代表者

堀内 匡 (HORIUCHI, Tadashi)

松江工業高等専門学校・電子制御工学科・教授

研究者番号：50294129

交付決定額(研究期間全体)：(直接経費) 3,600,000円

研究成果の概要(和文)：本研究では、実機の多脚ロボットを対象として、CPG (Central Pattern Generator) と強化学習を用いた枠組みにより、ロボット自身が障害物を回避しながら目標地点に到達するという複数タスクを遂行するための行動獲得を実現した。まず、それぞれのタスクを遂行する各行動モジュールをCPGと強化学習を用いて実現した。次に、ロボットの状況に応じて、行動モジュールの出力を統合する上位モジュールを導入した複数タスク環境での行動学習システムを構築した。六脚ロボットと四脚ロボットに加えて、ヘビ型ロボットにもこの行動学習システムを実装し、各ロボットにおける実機実験を通して、有効性を確認した。

研究成果の概要(英文)：In this research, we realized that real multi-legged robots themselves acquired behaviors to perform multiple tasks such as avoiding the obstacles and reaching the target area, by using Central Pattern Generator (CPG) and reinforcement learning. First, we realized each learning module to perform each task, by using CPG and reinforcement learning. Next, we constructed behavior learning system for multi-task environment, by adding the upper module to integrate the outputs of lower learning modules. We implemented this behavior learning system for six-legged robot, four-legged robot and snake-like robot. Through the experiments for three kinds of real robots, we confirmed the effectiveness of our behavior learning system using CPG and reinforcement.

研究分野：知能システム

キーワード：知能ロボティクス ニューラルネットワーク 強化学習 行動獲得 複数タスク

1. 研究開始当初の背景

強化学習は、ロボットなどの自律システムの学習手法としてだけでなく、未知の環境に対する最適化の手法としても有望である。しかし、問題の規模が大きくなると膨大な試行錯誤の回数を必要とする欠点がある。そのため、ロボットなどの高次元・実環境における応用は重要な課題である。一方、生物の運動を規範とした歩行パターンの生成機構としてCPG (Central Pattern Generator) と呼ばれる神経振動子モデルがある。環境に適応した歩行をロボットで実現するためにCPGを用いた歩行パターン生成に関する研究が行われてきた。しかし、CPGには環境やロボットに応じて決めるべき複数のパラメータがあり、実環境でのそれらの調整は難しいという問題点がある。

CPGのパラメータを強化学習により最適化する先駆的な研究としてCPG-actor-critic法がある。我々はこの手法に注目し、より簡略化した枠組みで実機の六脚ロボットや四脚ロボットが目標方向に向かう滑らかな動きを獲得できることを示した。具体的には、CPGには三角関数を利用したシンプルなCohenモデル、強化学習器にはQ学習を用いた。この枠組みでは、ロボットの全モータの目標角度を学習するのではなく、CPGの少数のパラメータを強化学習により学習・最適化する。そのため、問題の規模を抑えることができ、効率的な学習を実現した。

本研究では、新たな問題設定として、六脚ロボットや四脚ロボットなどの多脚ロボットを対象に、目標地点に到達する歩行動作を

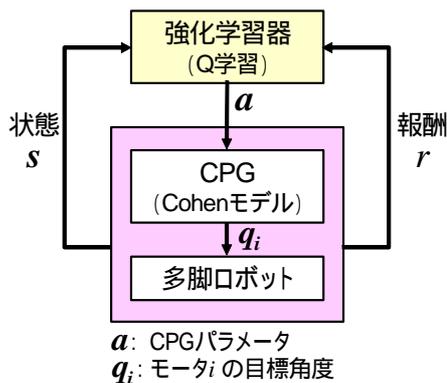


図1. 各行動モジュール内の枠組み

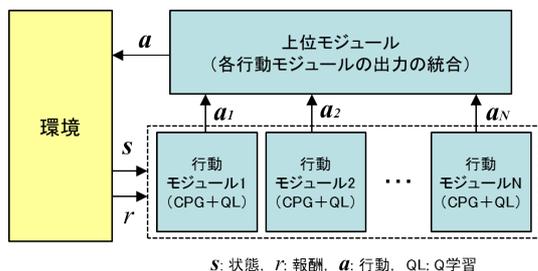


図2. マルチタスク環境における行動学習システム

学習するタスク、障害物を回避する歩行動作を学習するタスクなどの複数のタスクを考える。各タスクを遂行する各行動モジュールはCPGと強化学習を用いた枠組み(図1参照)で実現する。さらに、ロボットの状況に応じて、行動モジュールの出力を統合する上位モジュールを導入したマルチタスク環境での行動学習システムを構築する(図2参照)。具体的には、マルチタスクとして二つのタスクを想定し、多脚ロボットが障害物を避けながら目標地点に向かう行動の獲得を実現する。

2. 研究の目的

本研究では、六脚ロボットと四脚ロボットを対象として、以下に示す三つを研究の目的とする。

第一目的として、「CPGと強化学習を用いた目標到達の歩行動作獲得(行動モジュール1)」を実現する。CPGと強化学習を用いて目標地点に到達するための歩行動作の獲得を実現する。

第二目的として、「CPGと強化学習を用いた障害物回避の歩行動作獲得(行動モジュール2)」を実現する。CPGと強化学習を用いて障害物回避を行うための歩行動作の獲得を実現する。

第三目的として、「複数の行動モジュールの出力を統合する機能の実現(上位モジュール)」を実現する。目標到達タスクを遂行する行動モジュールと障害物回避タスクを遂行する行動モジュールの各出力を状況に応じて統合し、多脚ロボットが障害物を回避しつつ目標地点に到達するという全体タスクの実現を図る。

3. 研究の方法

まず、本研究の第一目的である「CPGと強化学習を用いた多脚ロボットの目標到達の歩行動作獲得(行動モジュール1)」の実現に関して、六脚ロボット(図3参照)と四脚ロボット(図4参照)を対象として、以下の(1)と(2)の方法により研究を実施した。

(1) CPGと強化学習を用いた目標到達行動の獲得(六脚ロボット)

六脚ロボットが目標地点に向けて向きを変えながら滑らかに進む歩行動作の獲得を実現する。

(2) CPGと強化学習を用いた目標到達行動の獲得(四脚ロボット)

四脚ロボットが目標地点に向けて向きを変えながら滑らかな歩行動作の獲得を実現する。

次に、本研究の第二目的である「CPGと強化学習を用いた多脚ロボットの障害物回避

の歩行動作獲得(行動モジュール2)の実現に関して、以下の(3)の方法により研究を実施した。

(3) CPG と強化学習を用いた障害物回避行動の獲得

六脚ロボットおよび四脚ロボットが障害物を回避するための歩行動作の獲得を実現する。

さらに、本研究の第三目的である「複数の行動モジュールの出力を統合する機能」の実現に関して、以下の(4)の方法により研究を実施した。

(4) 複数の行動モジュールの出力を統合する機能の実現

障害物回避の行動モジュールの出力と目標到達の行動モジュールの出力を上位モジュールで調停する機能を実現する。

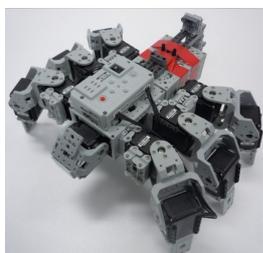


図 3. 六脚ロボット



図 4. 四脚ロボット

4. 研究成果

(1) CPG と強化学習を用いた目標到達行動の獲得(六脚ロボット)

六脚ロボットが目標地点に向けて向きを変えながら滑らかに進む歩行動作の獲得を実現した。その際、目標地点までの距離と角度が小さくなるほど報酬を大きく設定することで、目標地点に到達する行動を獲得することができた。

六脚ロボットの実験環境の模式図を図5に示す。六脚ロボットの本体上部に二色のマークを取り付け、天井近くに設置した Web カメラでマークを追跡するプログラムを実装し、ロボットの位置と方向を算出した。

六脚ロボットを用いた実験結果の一例を図6と図7に示す。学習初期ではロボットは目標領域に到達していないが、学習後にはスタートの方向によらず目標領域に到達する行動が獲得できていることがわかる。

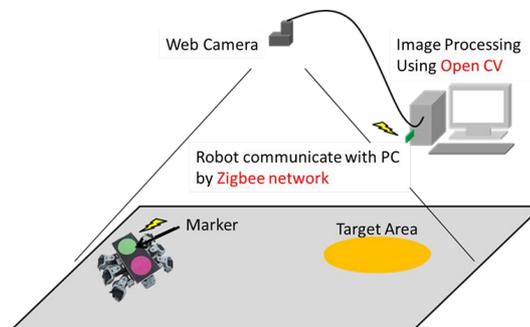


図 5. 六脚ロボットの実験環境 (目標到達タスクの学習)

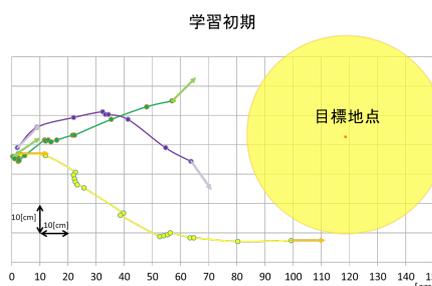


図 6. 学習初期における六脚ロボットの行動軌跡の例(目標到達タスク)

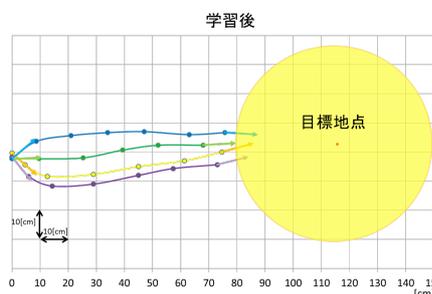


図 7. 学習後における六脚ロボットの行動軌跡の例(目標到達タスク)

(2) CPG と強化学習を用いた目標到達行動の獲得(四脚ロボット)

四脚ロボットが目標地点に向けて向きを変えながら滑らかな歩行動作の獲得を実現した。

四脚ロボットの実験環境の模式図を図8に示す。四脚ロボットにも本体上部に二色のマークを取り付け、天井近くに設置した Web カメラでマークを追跡し、ロボットの位置と方向を算出した。

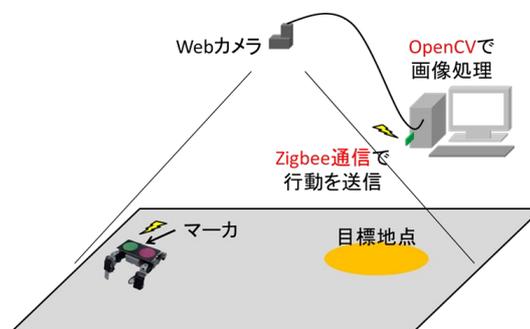


図 8. 四脚ロボットの実験環境 (目標到達タスクの学習)

四脚ロボットを用いた実験結果の一例を図9と図10に示す。四脚ロボットにおいても、学習初期ではロボットは目標領域に到達していないが、学習後にはスタートの方向によらず目標領域に到達する行動が獲得できていることがわかる。

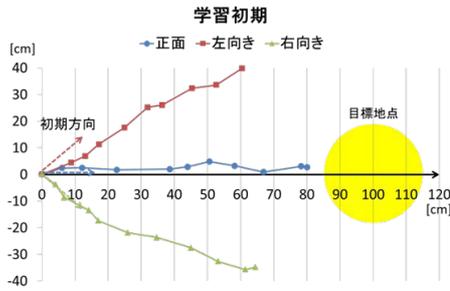


図9. 学習初期における四脚ロボットの行動軌跡の例（目標到達タスク）

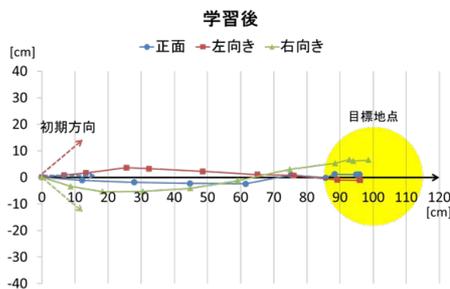


図10. 学習後における四脚ロボットの行動軌跡の例（目標到達タスク）

(3) CPG と強化学習を用いた障害物回避行動の獲得

六脚ロボットおよび四脚ロボットが障害物を回避するための歩行動作の獲得を実現した。障害物までの距離と角度が大きいほど報酬を大きく設定することで、障害物から離れる行動を獲得することができた。実機実験では、ロボットの本体に取り付けたマーカとは異なる色のマーカを障害物にも取り付け、実験室の天井近くに設置した Web カメラでロボットのマーカと障害物のマーカを追跡することにより、ロボットから見た障害物の位置と方向を算出した。

四脚ロボットを用いた場合の実験環境の模式図を図11に示す。

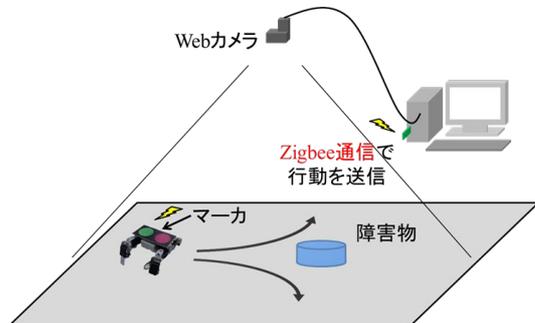


図11. 四脚ロボットの実験環境（障害物回避タスクの学習）

四脚ロボットを用いた実験結果の一例を図12に示す。障害物回避タスクの学習において、学習初期ではロボットは障害物領域に入ることが多いが、学習後にはスタートの方向によらず障害物を回避する行動が獲得できていることがわかる。

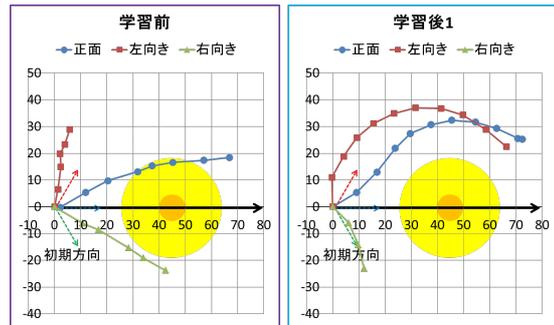


図12. 学習初期と学習後の四脚ロボットの行動軌跡の例（障害物回避タスク）

(4) 複数の行動モジュールの出力を統合する機能の実現

障害物回避の行動モジュールの出力と目標到達の行動モジュールの出力を上位モジュールで調停する機能を実現した。ロボットが障害物に接触する可能性がある状況では障害物回避行動の優先度を高くし、障害物が近くにない状況では目標到達行動の優先度を高くすることを検討した。

障害物回避の行動学習モジュールと目標到達の行動学習モジュールと上位モジュールから構成される行動学習システムの枠組みを図13に示す。

四脚ロボットを用いた場合の実験環境の模式図を図14に示す。

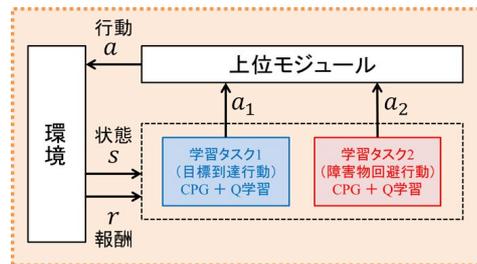


図13. 複数タスク（目標到達と障害物回避）環境における行動学習システム

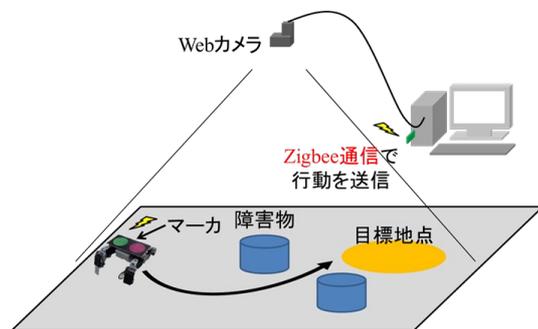
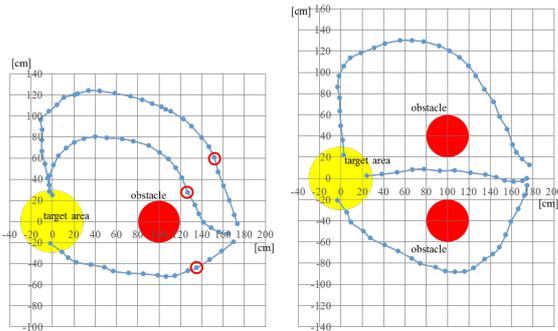


図14. 四脚ロボットの実験環境（複数タスク環境における行動学習）

四脚ロボットを対象とした実験結果の一例を図 15 に示す。学習初期ではロボットは障害物に衝突したり目標物の方向に向かわなかったりしていたが、学習後には障害物を回避しつつ目標物に到達する行動が獲得できていることがわかる。



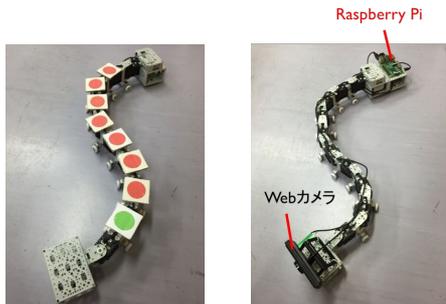
(a) 1つの障害物の場合 (b) 2つの障害物の場合

図 15. 学習後の四脚ロボットの行動軌跡の例 (複数タスク環境)

(5) CPG と強化学習を用いたヘビ型ロボットの行動獲得の実現

本研究の発展の方向性の一つとして、CPGで各モータを駆動するヘビ型ロボット(図 16 参照)を製作し、目標物に向けて進む行動を強化学習により獲得することを実現した。

実験結果の一例を図 17 に示す。学習初期ではロボットは目標物に向かわない行動ばかりであったが、50 試行の学習後には、スタートの向きに関わらず、目標物に到達する行動を獲得することができた。



a) ヘビ型ロボット (b) 自律型ロボット

図 16. ヘビ型ロボットの実機

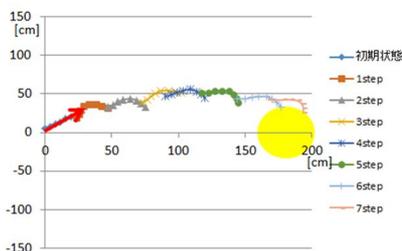


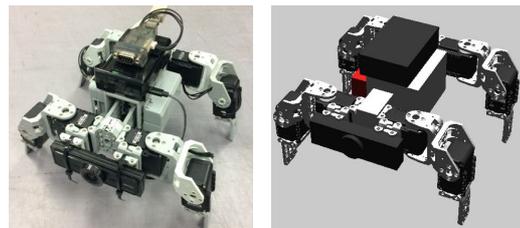
図 17. 学習後のヘビ型ロボットの行動軌跡の例 (ロボットのスタートの向きが目標物に対して斜め左の場合)

(6) 深層強化学習を用いた自律型四脚ロボットの行動獲得の実現

本研究の更なる発展として、小型の高性能コンピュータ Raspberry Pi と Web カメラを搭載した自律型四脚ロボットを製作し、深層強化学習を用いて、四脚ロボットが自分自身の視覚情報 (カメラ画像) を用いて、障害物を回避しながら目標物に到達する行動を獲得することを実現した。これにより、天井付近に Web カメラを設置する必要がなくなり、実験場所の制約がなくなった。

具体的には、深層強化学習の代表的な手法の一つである Deep Q-network (DQN) を用いて、ロボットに搭載したカメラ画像を状態として、左折 / 直進 / 右折 / 後退の 4 種類の行動の行動価値 (Q 値) を学習した。その際、四脚ロボットのシミュレーションモデル (図 18 参照) を構築し、そのモデルを利用してシミュレーション環境において学習した。シミュレーション環境で学習した DQN を用いて、実機環境でも継続して学習を進めた。DQN を用いた自律型四脚ロボットの行動学習の枠組みを図 19 に示す。

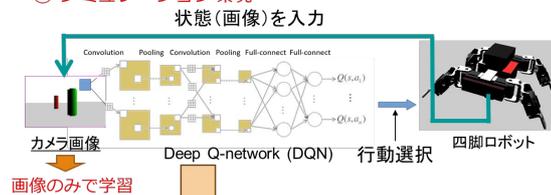
実験において、障害物は緑色の物体、目標物は赤色の物体とした。ロボットが障害物に衝突した際および目標物に到達した際には、ロボットが自分自身で方向を大きく変えて再スタートする自動リスタート機能を実装した。これにより、実機実験中に人間がロボットをスタート位置に置き直す手間がなくなり、実験を自動化することができた。



(a) 実機ロボット (b) モデル

図 18. 自律型四脚ロボットの実機とシミュレーションモデル

① シミュレーション環境



② 実機環境

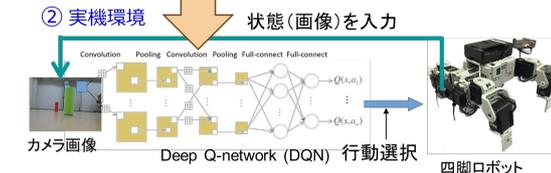


図 19. DQN を用いた自律型四脚ロボットの行動学習の枠組み

実験結果の一例を図 20 に示す。学習の進行とともに、目標物到達の回数が増え、障害物衝突の回数が減っており、自律型四脚ロボットが視覚情報を用いて適切な行動を獲得していることがわかる。

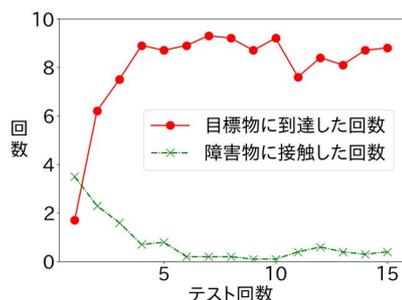


図 20. 学習の進行の様子(各テストにおける目標物到達と障害物衝突の回数の推移)

5. 主な発表論文等

〔雑誌論文〕(計 6 件)

H. Sasaki, T. Horiuchi and S. Kato, Experimental Study on Behavior Acquisition of Mobile Robot by Deep Q-network, Journal of Advanced Computational Intelligence and Intelligent Informatics, 査読有, Vol.21, No.5, pp.840-848, 2017
DOI: 10.20965/jaciii.2017.p0840

H. Sasaki, T. Horiuchi and S. Kato, A Study on Behavior Acquisition of Mobile Robot by Deep Q-network, ICIC Express Letters, Part B: Applications, 査読有, Vol.8, pp. 727-733, 2017
DOI: 10.24507/icicelb.08.04.727

永海 昂, 岸本良一, 堀内 匡, CPG と強化学習を用いたヘビ型ロボットの目標到達行動の獲得, 知能と情報(日本知能情報ファジィ学会誌), 査読有, Vol.29, No.2, pp.551-557, 2017
DOI: 10.3156/jsoft.29.2_551

石倉裕貴, 岸本良一, 堀内 匡, CPG と強化学習を用いた多脚ロボットの目標到達行動の獲得, 電気学会論文誌 C(電子・情報・システム部門誌), 査読有, Vol.136, No.3, pp.333-339, 2016
DOI: 10.1541/ieejieiss.136.333

〔学会発表〕(計 28 件)

加藤大地, 佐々木 光, 堀内 匡, 青代敏行, 深層強化学習の実機ロボットへの応用: 車輪型移動ロボットの行動獲得の実現, 計測自動制御学会 第 45 回知能システムシンポジウム, 2018

吉木綜一郎, 田中 駿, 堀内 匡, 幸田憲明, 深層強化学習を用いた四脚ロボットの視覚ベースの行動獲得の実現, 電気学会 システム研究会, 2018

H. Sasaki, T. Horiuchi and S. Kato, A Study on Vision-based Mobile Robot

Learning by Deep Q-network, The SICE Annual Conference 2017, 2017

H. Sasaki, T. Horiuchi and S. Kato, Vision-based Behavior Learning of Mobile Robot by Deep Q-network, The 12th International Conference on Innovative Computing, Information and Control, 2017

佐々木 光, 堀内 匡, 加藤 聡, 深層強化学習を用いた移動ロボットの視覚ベースの行動学習, 平成 29 年電気学会電子・情報・システム部門大会, 2017

吉木綜一郎, 堀内 匡, 幸田憲明, CPG と強化学習を用いた自律型四脚ロボットの行動獲得の高速化, 平成 29 年電気学会電子・情報・システム部門大会, 2017

堀内 匡, 機械学習の実機ロボットへの応用, 計測自動制御学会 第 6 回知能工学部会研究会「賢さの先端研究会」/ 第 54 回システム工学部会研究会, 2016

佐々木 光, 堀内 匡, 加藤 聡, 深層強化学習を用いた移動ロボットの行動獲得の試み, 平成 28 年電気学会電子・情報・システム部門大会, 2016

吉木綜一郎, 三上大志, 堀内 匡, CPG と強化学習を用いた自律型四脚ロボットの行動獲得, 平成 28 年電気学会電子・情報・システム部門大会, 2016

三上大志, 永海 昂, 岸本良一, 堀内 匡, CPG と強化学習を用いた自律型四脚ロボットの行動獲得の試み, 平成 28 年電気学会全国大会, 2016

永海 昂, 岸本良一, 堀内 匡, CPG と強化学習を用いたヘビ型ロボットの目標到達行動の自律的獲得, 電気学会 システム研究会, 2016

岸本良一, 石倉裕貴, 堀内 匡, CPG と強化学習を用いた四脚ロボットの複数タスクの獲得, 平成 27 年電気学会電子・情報・システム部門大会, 2015

永海 昂, 岸本良一, 堀内 匡, CPG と強化学習を用いたヘビ型ロボットの行動獲得の試み, 平成 27 年電気学会電子・情報・システム部門大会, 2015

佐々木 光, 堀内 匡, 加藤 聡, 畳み込みニューラルネットワークの古文書文字認識への応用, 平成 27 年電気学会電子・情報・システム部門大会, 2015

6. 研究組織

(1) 研究代表者

堀内 匡 (HORIUCHI, Tadashi)
松江工業高等専門学校・電子制御工学科・教授
研究者番号: 50294129

(2) 研究協力者

岸本 良一 (KISHIMOTO, Ryoichi)
佐々木 光 (SASAKI, Hikaru)