

接触状態を含む身体モデルと強化学習を用いた筋骨格ヒューマノイドによる環境接触行動

Realization of Multi Contact Behaviors Using Contact-State Transition Model and Reinforcement Learning by Musculoskeletal Humanoid

西浦 学 (東大) 正 河原塚 健人 (東大)
学 利光 泰徳 (東大) 正 浅野 悠紀 (東大)
岡田 慧 (東大) 川崎 宏治 (トヨタ自動車)
稻葉 雅幸 (東大)

Manabu NISHIURA, The University of Tokyo, nishiura@jsk.imi.i.u-tokyo.ac.jp
Kento KAWAHARAZUKA, Yasunori TOSHIMITSU, Yuki ASANO, Kei OKADA, The University of Tokyo
Koji KAWASAKI, TOYOTA MOTOR CORPORATION, DENSO CORPORATION SOKEN, inc
Masayuki INABA, The University of Tokyo

A human-being often does contact-rich behaviors in daily lives. Contact-rich behaviors require human or robots to handle friction force between environment and their body. We propose contact-state transition model to achieve contact-rich behaviors. The contact-state transition model represents dynamics of a robot and an environment which contacts with the robot. First, we confirm the effectiveness of contact-state transition model by conducting the hip shuffle experiment on the chair. Finally, we combine contact-state transition model and reinforcement learning, and realize sit back motion in the musculoskeletal humanoid.

Key Words: Tactile Sensor, Reinforcement Learning, Musculoskeletal Model

1 はじめに

人間が日常生活で行う行動のうち、多くは身体の多点を環境に接触させて行う環境接触行動と呼ばれる。環境接触行動に適したヒューマノイドロボットとして人体を模倣した筋骨格ヒューマノイドがある[1], [2]。筋骨格ヒューマノイドは人体同様やわらかい身体を持ち、環境に柔軟に身体をなじませながら課題をこなすことができる。また、弾性体からなる筋アクチュエータを全身に持つことで弾性をいかして衝撃を吸収することができること、冗長な筋配置により筋が一本切れても動き続けられることなどの利点がある。一方、筋の経由点での摩擦や動作中の筋経路変化、異なる素材を用いたやわらかい身体のモデル化が困難であることなどの理由により正確に動かすことが難しく、機械学習を用いた制御手法などが提案されている[3][4]。

筋骨格ヒューマノイドにおいて環境接触行動を実現した例としては[5]などがある。[5]においては、ペダル操作角度と車両速度の関係を表す非線形なモデルをニューラルネットワークで学習し、学習したモデルを用いてモデル予測制御を行うことでペダル操作を実現している。

また、ロボットの実機とシミュレーションではダイナミクスに差異があることが知られている[6]。特にロボットと環境物体との摩擦はダイナミクスの差異に大きく寄与する。そのため、環境物体とロボット身体が接触するなかで課題をこなす必要がある環境接触行動をシミュレーションで学習し、学習したモデルをそのまま用いて実機で課題を遂行することは困難であると考えられる。

本研究では多点環境接触行動を実現するために、課題を実行するための目標状態列をシミュレーションで学習し、目標状態間を遷移するための接触状態を含んだ身体モデルを実機で学習し、組み合わせることで多点環境接触行動を実現する方法を提案する。多点環境接触行動の一例として、椅子への座り込み、座り替え動作を扱う。

2 多点環境接触行動の学習の概要

本研究では手先足先以外の身体部位での環境接触を含む行動を多点環境接触行動と定義する。図1に提案する学習システムの概要を示す。多点環境接触行動に対して、その動作を実行可能なロ

ボットのコンフィギュレーションと接触状態列をシミュレーションで強化学習を用いて探索する。加えて、実機を用いてロボットの身体の接触状態遷移を表現する身体モデルを自己教師あり学習を用いて学習する。シミュレーションで学習した目標状態列間を接触状態を含む身体モデルを用いて遷移することでシミュレーションと実環境のギャップを吸収しながら環境接触行動を行うようなシステムになっている。

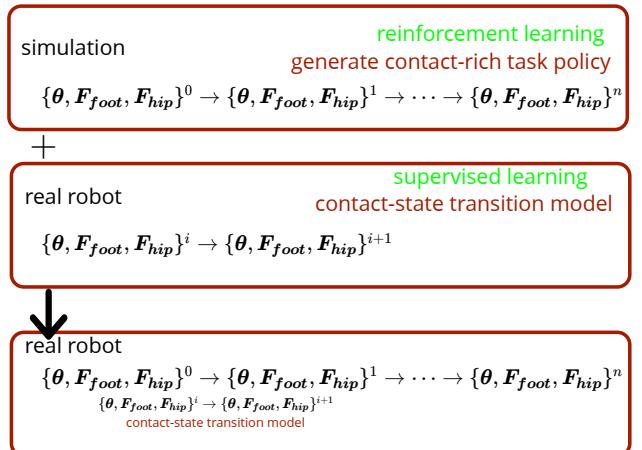


Fig.1 Overview of learning system of contact-rich behaviors using contact-state transition.

3 接触状態を含んだロボットの状態遷移モデル

3.1 Contact-State Transition Model

図2に接触状態を含んだロボットの状態遷移モデルの概要を示す。ヒートマップは臀部と足先の接触センサにかかる力を可視化したものである。 s_t は現在の状態を表し、 a_t は現在の状態における制御入力を表し、 s_{t+1} は1タイムステップ先の状態を表

す。状態方程式をニューラルネットワークで表現する。

$$s_{t+1} = f(s_t, a_t) \quad (1)$$

が図 2 に相当する。状態方程式をニューラルネットワークで学習することにより、目標状態を s_{ref} 、求める制御入力を a_{opt} として、学習した状態方程式からの出力が目標状態 s_{ref} に近づくように制御入力 a_{opt} を更新することにより目標状態を実現する制御入力を得る手法が提案されている [7][8][9]。更新則を以下に示す(式 4)。

$$L = h_{loss}(s^{est}, s^{ref}) \quad (2)$$

$$g = \partial L / \partial a^{opt} \quad (3)$$

$$a^{opt} \leftarrow a^{opt} - \alpha g / \|g\|_2 \quad (4)$$

これは、勾配法を用いた最適制御とみなすこともできる。

提案する接触状態を含んだ状態遷移モデルでは、以下のように状態を表現する。

$$s = \{F_{footloadcell}, F_{hipfsr}, \theta\} \quad (5)$$

$$a = \{\Delta\theta\} \quad (6)$$

本研究では筋骨格ヒューマノイド MusashiOLEgs[10] を用いて椅子に座っている状態での接触状態遷移モデルを学習する。 $F_{footloadcell}$ は左右それぞれで 4 つの足裏ロードセルの値を用いるため、8 次元のベクトル、 F_{hipfsr} も左右それぞれで 4 つの fsr センサの値を用いるため、8 次元のベクトルである。 θ は座った状態での重心移動に関わる {spine-roll, lleg-hip-pitch, lleg-hip-roll, rleg-hip-pitch, rleg-hip-roll} 軸の 5 次元のベクトルとした。

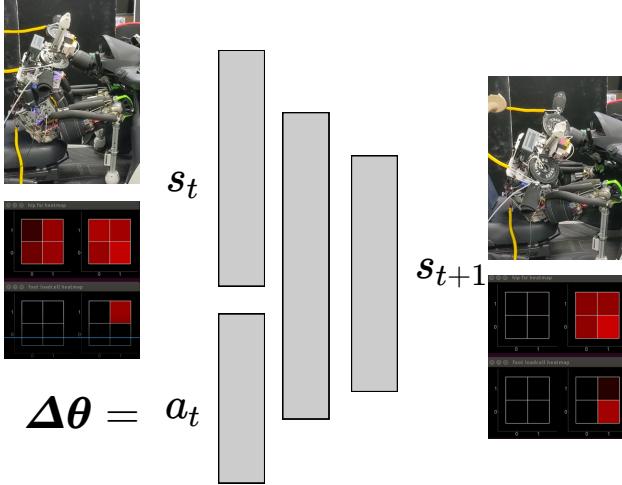


Fig.2 Overview of contact-state transition model.

3.2 接触状態を含んだ状態遷移モデルの学習

本研究では筋骨格ヒューマノイド MusashiOLEgs[10] を用いて椅子に座っている状態での接触状態遷移モデルを学習した。図 3 に接触状態を含んだ状態遷移モデルを学習するためのデータを集めている様子を示す。図 3 下の図は上の姿勢のときの臀部触覚センサを値を可視化した図である。様々な状態 s_t において制御入力 a_t を与えたときの状態変化 s_{t+1} を記録している。MusashiOLEgs を用いて椅子に座った状態で外部 PC から目標関節角度を送り自己身体像モデル [3] を用いて目標関節角度を実現するような筋長を出して筋モジュールに送っている。

図 3 左下では重心が左によっているために、左の臀部にとりつけられた触覚センサの値が大きく変化していることがわかる。図 3 右下では姿勢が変化し、重心が右によっているために、右の臀部にとりつけられた触覚センサの値が大きく変化していることがわかる。

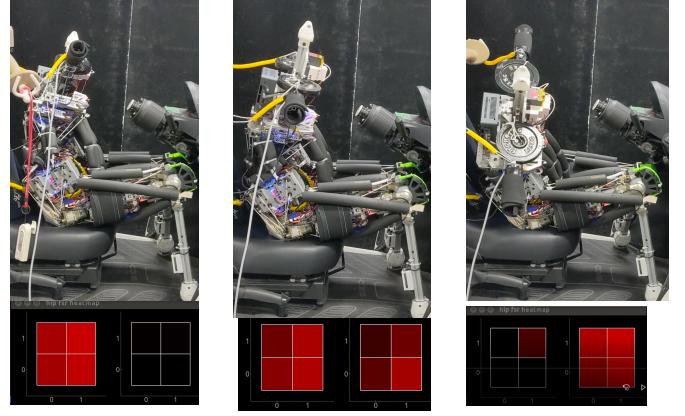


Fig.3 MusashiOLEgs collecting data to learn contact-balance model.

集めたデータを用いて状態遷移モデルを学習した。状態遷移モデルは全結合層-NStepLSTM-全結合層からなり、損失関数には平均二乗誤差 (MSE: Mean Squared Error) を用い、更新則には Adam[11] を用いた。全結合層の出力には Batch Normalization[12] を用いている。

4 強化学習による目標状態生成

4.1 問題設定

強化学習は以下の 5 つ組 $M \triangleq \{\mathcal{S}, \mathcal{A}, p_{s_0}, p_T, g\}$ で定義されるマルコフ決定過程に対して考えられる [13]。

- 有限状態集合: $S \triangleq \{1, \dots, |S|\} \ni s$
- 有限行動集合: $A \triangleq \{a^1, \dots, a^{|\mathcal{A}|}\} \ni a$
- 初期状態確率関数: $p_{s_0} : \mathcal{S} \rightarrow [0, 1]$
- 状態遷移確率: $p_T : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$
- 報酬関数: $g : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$

マルコフ決定過程上で定義される確率的方策 $\pi : \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$ を期待割引累積報酬 (式 7) を最大化するようにパラメータを更新していく。

$$\mathbb{E}^\pi [\sum_{t=1}^T \gamma^{t-1} g(S_t, A_t)] \quad (7)$$

観測 s_t は

$$s_t = \{F_{foot}, F_{hip}, \theta, \theta_z, \theta_x, \theta_y\} \quad (8)$$

とし、椅子への座り込み課題の報酬関数は (式 9) とした。

$$r = c_1 r_{footcontact} + c_2 r_{hipcontact} - r_{footdistance} + r_{onchair} \quad (9)$$

$r_{footcontact}, r_{hipcontact}$ は足のロードセルと臀部触覚センサをそれぞれ正規化した値、 $r_{footdistance}$ は椅子の正面に座るために設定したそれぞれの足先の目標座標と現在の足先の距離の L2 ノルム、 $r_{onchair}$ は毎エピソード毎に椅子から落ちていなければ加算される定数報酬となっている。また、 c_1, c_2 はそれぞれ 0.1、定数報酬は 0.2 とした。図 5 に $r_{footdistance}$ に用いた目標足先座標を示す。行動 a_t は

$$a_t = \{\Delta\theta\} \quad (10)$$

とした。物理シミュレータ内でも実機同様自己身体像モデル [3] を用いて目標関節角度を実現するような筋長を出力して筋を駆動することでロボットを動かしている。

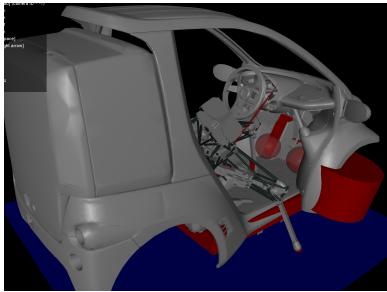


Fig.4 Simulation setting and target foot position.

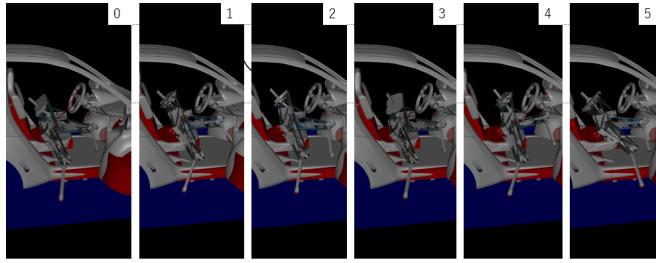


Fig.5 Sitting motion learned in simulation.

4.2 座り込み動作の学習

方策の学習アルゴリズムには Soft-Actor-Critic 法 [14] を用いた。物理シミュレータには MuJoCo[15] を用いた。図 5 に学習した方策による椅子への座り込み動作を示す。足先でうまく力を発揮して自身の姿勢を変化させているところが確認できる。

5 実験

本研究で提案した接触状態を含む身体モデルを用いた実験を 2 種類行う。1 つめは接触状態を含む身体モデルの評価実験で、椅子に座った状態から深く座り込む動作を行う。2 つめは椅子に腰かけた状態から正面に向かって座り直し動作を行う。

5.1 筋骨格ヒューマノイドによる椅子への座り込み行動

本節では接触状態を含んだ身体モデルの評価実験のため、椅子に正面から座った状態から深く座り込む動作を行う。ロボットは重心を左右に移動させ、臀部の片側のみに重心を移動し、重心がかかる方向の臀部とその側の脚を軸にしてもう片側の臀部を後ろに送ることが要求される。座り込み動作時の概要を図 6 に示す。座り込み動作を 3 つの段階に分け、それらの繰り返しにより実現する。動作は重心移動フェーズ、臀部移動フェーズ、リセットフェーズの 3 段階からなる。椅子に浅く腰かけた状態を初期姿勢とする。3 つのフェーズを左臀部を椅子の後ろに送る場合の例で示す。重心移動フェーズでは 3 で学習した状態遷移モデルを用いて右臀部に重心を移動する。臀部移動フェーズでは右臀部と右足先を軸に股関節ロール軸を回転させることで左臀部を椅子後ろに送る。リセットフェーズでは初期姿勢に姿勢を変更して次の繰り返しに備える。

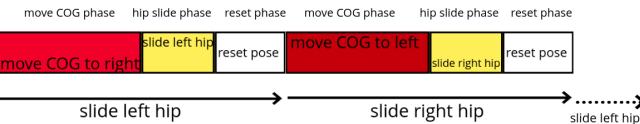


Fig.6 Overview of hip-slide motion using contact-state transition model.

図 7 に座り込み動作の様子を示す。図 7 下に座り込み動作時の臀部触覚センサにかかる力を示す。左臀部を椅子の後ろに送るときは右臀部に重心を寄せることで左臀部と椅子の摩擦を減らすことができていることが確認できる。

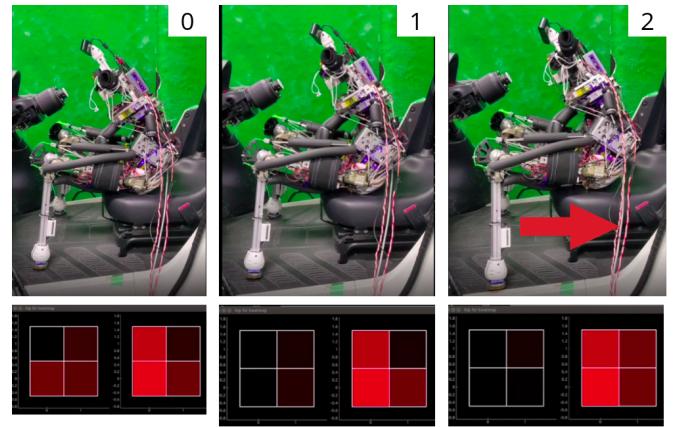


Fig.7 MusashiOLEgs sits back using contact-state transition model.

5.2 筋骨格ヒューマノイドによる椅子への座り直し行動

本節では 2 節で説明したシステムを用いて椅子への座り込み動作を行う。本研究では車両への乗り込みを想定し、自動車の椅子に横から腰かけた状態から正面に向かって座り直す動作を扱う。自動車にはトヨタ自動車製の超小型 EV 「コムス」シリーズの B-COM デリバリーを使用した。

図 8 に座り直し動作の概要を示す。4 節で探索した目標状態列のうち関節角度列のみを用いて初期姿勢 s_0 から s_1 に遷移していくが、シミュレーションと実機のダイナミクスの差異や初期セットアップの違いなどにより現実には s'_1 に遷移する。そこで 3 節で学習した接触状態を含む状態遷移モデル f を用いて s_1 に近づくような制御入力を用いて行動を行う。

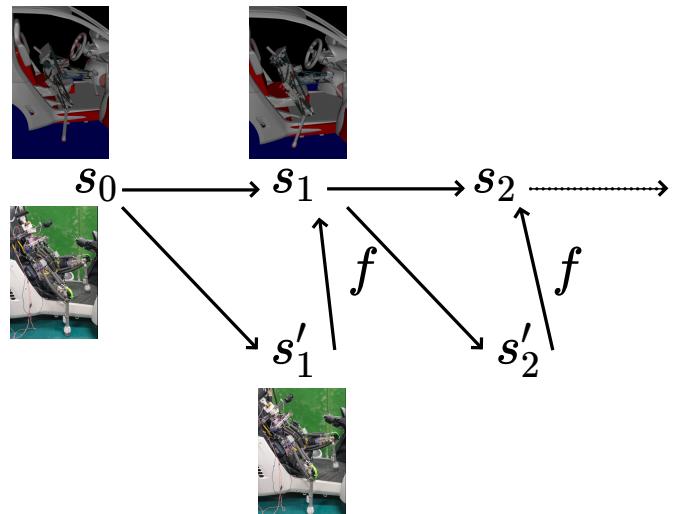


Fig.8 Task execution using contact-state transition model and reinforcement learning.

図 9 に実験時のように座り直す動作を示す。椅子に腰かけた状態を初期状態として車両の正面に向かって座った状態を目標状態とする。

重心を変化させつつ足先で力を発揮することで椅子に腰かけた状態から車両正面に向かって姿勢を変更することができていることが確認できる。

6 考察

本研究では環境接触行動に適したロボットとして筋骨格ヒューマノイドを用い、動作中に環境物体とロボット身体の間で広い面

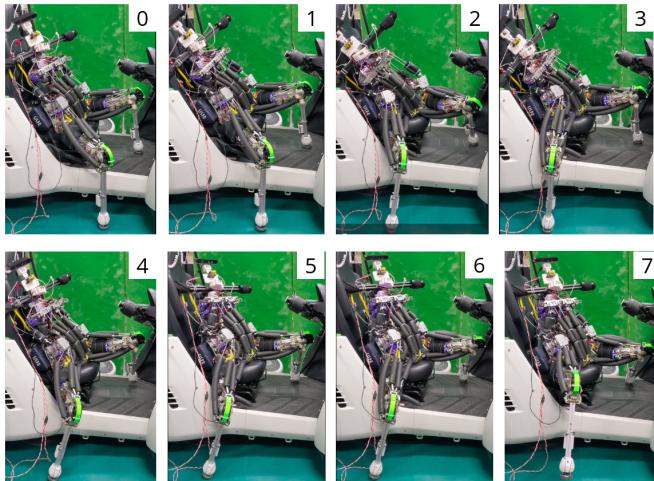


Fig.9 MusashiOLEgs gets in the car.

積で摩擦が発生するような課題を扱った。高い減速比のモーターによる軸駆動ロボットでこのような動作を行う場合、関節負荷などを動作探索の際に制約条件に入れる必要がある。[16]では臀部リンクの軌道を与えた状態で関節負荷などを制約条件に加えて、人間が与えた接触状態を遷移可能な軌道を探査することで座り込みを実現している。本研究では軌道を与えることなく、課題に関する情報である報酬関数からロボットが自身でシミュレータ内で学習することで座り込み動作を実現している。また、[16]では基本的に足裏を地面に固定し、地面と鉛直な平面上を臀部リンクが2次元的に移動するが、本研究では人間の座り込み動作を参考に3次元的に臀部リンクを動かしながら座り込む動作を可能にする学習システムの提案を行った。

接触状態を含んだ状態遷移モデルはロボットの身体とそれに対する環境物体を含めたダイナミクスの状態方程式を表現している。今回のように椅子に座った状態では主にロボットの重心変化を表すモデルになっている。筋骨格ヒューマノイド以外のロボットにおいても環境物体との接点に接触センサを取り付けなければ本研究と同様の状態遷移モデルを用いた非線形最適制御が行えると考える。

7 結論

本研究では、人間の日常生活に多く含まれる環境接触行動のうち、特に手先足先以外の身体部位での環境接触を含む動作を扱い、動作の軌道情報を人が与えることなくロボットが探索して獲得する手法について提案を行った。接触状態を含めて動作計画を行うため、シミュレーションで課題を実行可能な目標状態列を学習し、実機で学習した状態遷移モデルと組み合わせることで環境接触行動を実現した。本研究では、センサがある部位のみの状態遷移を考える必要がある動作を扱った。今後は腕と脚を同時に使う必要がある動作などに拡張していく必要がある。また、本研究では接触状態を含む身体モデルに関して椅子に座っている状態での状態遷移のみのモデルを学習した。今後は接触している環境物体を識別できるようなモデルなどに拡張していく必要がある。

8 謝辞

本研究の一部はトヨタ自動車株式会社の支援を受けて行われた。

参考文献

- [1] Asano, Y., Okada, K., and Inaba, M., “Design principles of a human mimetic humanoid: Humanoid platform to study human intelligence and internal body system”, *Science Robotics*, vol. 2-13, 2017.
- [2] Mizuuchi, I., Yoshikai, T., Sodeyama, Y., Nakanishi, Y., Miyadera, A., Yamamoto, T., Niemela, T., Hayashi, M., Urata, J., Namiki, Y., Nishino, T., and Inaba, M., “Development of musculoskeletal humanoid Kotaro”, In *Proceedings 2006 IEEE International Conference on Robotics and Automation*, 2006. ICRA 2006., pp. 82–87, 2006.
- [3] Kawaharazuka, K., Tsuzuki, K., Makino, S., Onitsuka, M., Asano, Y., Okada, K., Kawasaki, K., and Inaba, M., “Long-Time Self-Body Image Acquisition and Its Application to the Control of Musculoskeletal Structures”, *IEEE Robotics and Automation Letters*, vol. 4-3, pp. 2965–2972, 2019.
- [4] Kei, T., Kento, K., Tasuku, M., Moritaka, O., Shogo, M., Yuki, A., Kei, O., Koji, K., and Masayuki, I., “可動眼球と自己身体を用いた距離認識機能の獲得”, 日本機械学会ロボティクス・メカトロニクス講演会’19 講演論文集, 2019.
- [5] Kawaharazuka, K., Tsuzuki, K., Koga, Y., Omura, Y., Makabe, T., Shinjo, K., Onitsuka, M., Nagamatsu, Y., Asano, Y., Okada, K., Kawasaki, K., and Inaba, M., “Toward Autonomous Driving by Musculoskeletal Humanoids: A Study of Developed Hardware and Learning-Based Software”, *IEEE Robotics Automation Magazine*, vol. 27-3, pp. 84–96, 2020.
- [6] Tan, J., Zhang, T., Coumans, E., Iscen, A., Bai, Y., Hafner, D., Bohez, S., and Vanhoucke, V., “Sim-to-Real: Learning Agile Locomotion For Quadruped Robots”, In *Proceedings of Robotics: Science and Systems*, Pittsburgh, Pennsylvania, June 2018.
- [7] Kawaharazuka, K., Ogawa, T., Tamura, J., and Nabeshima, C., “Dynamic Manipulation of Flexible Objects with Torque Sequence Using a Deep Neural Network”, In *2019 International Conference on Robotics and Automation (ICRA)*, pp. 2139–2145, 2019.
- [8] Tanaka, D., Arnold, S., and Yamazaki, K., “EMD Net: An Encode-Manipulate-Decode Network for Cloth Manipulation”, *IEEE Robotics and Automation Letters*, vol. 3-3, pp. 1771–1778, 2018.
- [9] Kawaharazuka, K., Tsuzuki, K., Onitsuka, M., Asano, Y., Okada, K., Kawasaki, K., and Inaba, M., “Object Recognition, Dynamic Contact Simulation, Detection, and Control of the Flexible Musculoskeletal Hand Using a Recurrent Neural Network With Parametric Bias”, *IEEE Robotics and Automation Letters*, vol. 5-3, pp. 4580–4587, 2020.
- [10] Onitsuka, M., Nishiura, M., Kawaharazuka, K., Tsuzuki, K., Toshimitsu, Y., Omura, Y., Asano, Y., Okada, K., Kawasaki, K., and Inaba, M., “Development of Musculoskeletal Legs with Planar Interskeletal Structures to Realize Human Comparable Moving Function”, In *The 38th Annual Conference on Robotics Society of Japan*, pp. 2G2–08, 2020.
- [11] Kingma, D. P. and Ba, J. L., “Adam: A method for stochastic optimization”, In *3rd International Conference on Learning Representations*, 2015.
- [12] Ioffe, S. and Szegedy, C., “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift”, In *Proceedings of the 32nd International Conference on Machine Learning*, pp. 448–456, 2015.
- [13] D.P.Bertsekas, , *Dynamics Programming*, Athena Scientific, 1995.
- [14] Haarnoja, T., Zhou, A., Abbeel, P., and Levine, S., “Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor”, *arXiv e-prints*, p. arXiv:1801.01290, January 2018.
- [15] Todorov, E., Erez, T., and Tassa, Y., “MuJoCo: A physics engine for model-based control”, In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5026–5033, 2012.
- [16] 野田晋太郎, 野沢峻一, 垣内洋平, 岡田慧, 稲葉雅幸, “複数の接触遷移方式を統合する全身行動計画法とヒューマノイドの滑り接触遷移行動への応用”, 日本ロボット学会誌, vol. 35-5, pp. 393–402, 2017.