

# 深層学習によるロボット知能化のためのモーションキャプチャ教示システム

大山 知理 †‡      鈴木 彼方 †‡  
 † 早稲田大学理工学術院

陽 品駒 ‡      尾形 哲也 †‡  
 ‡ 産業技術総合研究所

## 1 はじめに

近年、生活の様々なシーンへのヒューマノイドロボットの導入が取り組まれている。実環境は人間の行動に合わせてできていることから、人間の身体に近い構成を持つロボットの活躍が期待されているためである。しかし、逐次的に変化する環境下で動作を行うことは困難なため、機械学習によるロボットの自律化が求められている。

従来研究では、多自由度ロボットの強化学習等の教示なし学習に基づく動作生成が行われている [1]。しかし、適切な動作を探索するには膨大な時間がかかり、実機による繰り返しの試行はロボットハードの疲弊につながる可能性がある。一方、教示あり学習による手法では、人がロボットを直接動かすダイレクトティーチングによる動作教示が広く用いられており、野田らも小型ロボットにダイレクトティーチングで動作を教示している [2]。しかし、この操作方法は複雑な動作教示が難しく、工業用等の巨大なロボットを扱う際に危険が伴う。そこで、陽・鈴木らは3D マウスや打ち込みによる遠隔操作で教示データを生成していたが [3]、両手の協調を必要とするような複雑な操作が難しく、片手で行うタスクを教示しているという問題があった。

本研究では、教示データ生成に直感的に操作可能なモーションキャプチャを用いることで動作教示を効率化し、両手の協調を必要とするような複雑な動作生成を可能とする。また、複数の深層学習器による学習を行い、多自由度ロボットに適用した。

## 2 モーションキャプチャによるロボットの遠隔操作

本研究で用いる動作教示システムを図1に示す。ロボット視点からの画像を取得・確認しながら、モーションキャプチャによる遠隔操作を行う。モーションキャプチャは、両手を使うなどの複雑な動作教示が容易であり、3D マウスなどのコントローラに比べ直感的な動作生成が可能となる。また、ロボット視点からの画像を確認しながらリアルタイムで操作を行うことで、変化する状況に対応しながら動作を生成することができる。

しかし、人間とロボットは自由度や可動域などの身体性に違いがあるため、人間のモーションを単純な変換でロボットの動作にすることはできない。そこで、タ

スク動作で必要となる手先位置と手先角度が取得できるように、モーション開始時と現在の手先位置・角度の変位から、ロボットの関節角度を逆運動学で計算し指定・実行した。

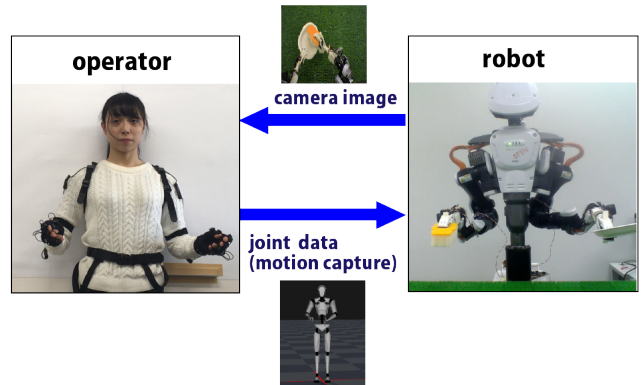


図 1: モーションキャプチャによるロボット操作の概要

## 3 深層学習モデル

本研究で用いる学習モデルを図2に示す。本研究では、二つの深層学習器を用いる。はじめに、タスク実行時のロボット視点のカメラ画像を畳みこみ層を有した Deep Convolutional Auto-Encode (DCAE) [4] により学習し、低次元の画像特徴量を得る (a)。得られた画像特徴量とロボットの動作データを合わせ、Recurrent Neural Network (RNN) を用いて運動情報と画像情報の関係性を学習する (b)。また、本研究では時系列データのプリミティブを考慮し、次の状態を予測することができる Multiple Timescales Recurrent Neural Network (MTRNN) [5] を用いることで、ロボットの現在の運動情報と画像情報から次の状態を予測することが可能となる。

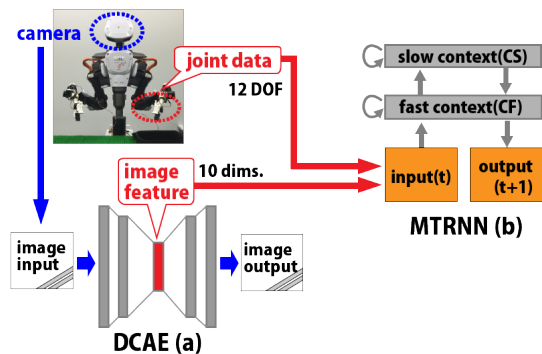


図 2: 学習モデル

Teaching System utilizing Motion Capture for Intelligence Robot with Deep Learning  
 Chisato Ohyama(Waseda Univ.), Kanata Suzuki(Waseda Univ.), Pin-Chu YANG(AIST), Tetsuya Ogata(Waseda Univ./AIST)

## 4 実験

### 4.1 実験設定

本研究では、実ロボットにおける動作生成として各腕 6 自由度のヒューノイドロボット Nextage[6] を用い、タスク動作を学習させた。また、モーションキャプチャには、3 軸ジャイロスコープ・3 軸加速度計・3 軸磁力計を搭載した Perception Neuron[7] を用いた。構築したシステムは、ロボットミドルウェアである ROS を用い、キャプチャした人の関節データをロボットの関節データに変換するモジュールと、Nextage のコントローラからなる。

タスク動作として、両腕の協調を必要とする皿洗いを行った。教示データとして、右手でスポンジ・左手で皿を持った状態から、皿を傾けスポンジで 5 回こすり、元の位置に戻るといった動作を、21 セット行った。各データは約 0.2step/seq で 310step (約 62 秒) 取得し、 $112 \times 112 \times 3$  ピクセルの RGB 画像から DCAE で 10 次元の画像特徴量に圧縮した。RNN の入力データとして、取得した画像特徴量 10 次元と両腕 12 自由度の、合計 22 次元を扱った。

### 4.2 結果と考察

モーションキャプチャでロボットを操作し動作生成した様子を図 3 に示す。また、動作生成時のロボットの各腕の関節角度の遷移を図 4 に示す。図中の実線は教示データ、点線は出力データを表している。動作生成した関節角度の二乗誤差の平均は 0.0295 であり、教示データと同じようにスポンジと皿を持った状態から、皿洗いの動作を生成することが可能となったと言える。モーションキャプチャで操作性を高めると実際に取得するロボットの動作には手の震えなどのノイズがかかるが、データセットを大量に取ることで MTRNN である程度動作を汎化し、問題なく動作を生成することが可能である。したがって、ロボット操作に慣れていない操縦者でも短時間で機械学習のための学習データを作成可能であると言える。しかしながら、操縦者へのフィードバックがないため、環境との接触時の安全性が課題として残る。

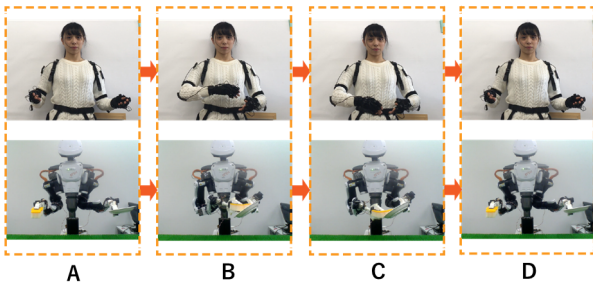


図 3: モーションキャプチャによるロボット操作と動作生成

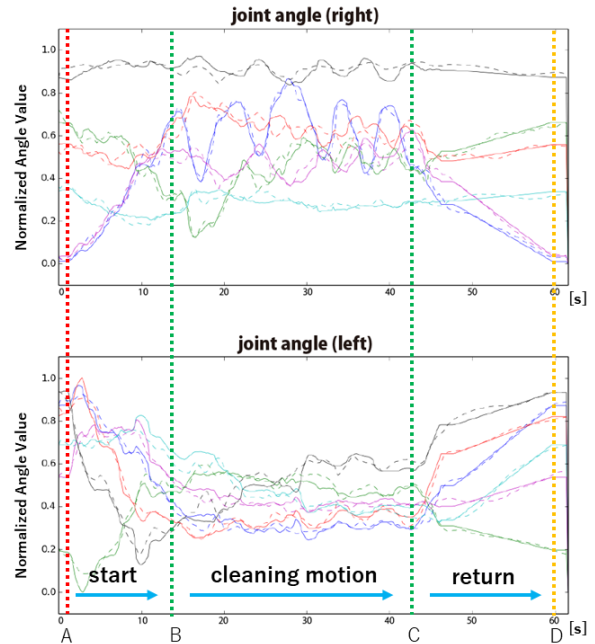


図 4: 動作生成結果 (関節角度)

## 5 おわりに

本研究では、ロボットの動作データ作成にモーションキャプチャを用いることで、効率的な動作教示システムを構築した。取得した教示データを深層学習器によって学習することで、実ロボットによる動作生成が可能となった。今後は操縦者へのフィードバックの追加や、より幅広いロボットハードへの適用を試みる。

## 謝辞

本研究は、文科省科研費基盤研究 (A)(No.15H01710), NEDO の支援の下で行われました。ここに謝意を表します。

## 参考文献

- [1] Levine Sergey, et al., “End-to-End Training of Deep Visuomotor Policies,” arXiv preprint arXiv:1504.00702 (2015)
- [2] Noda Kuniaki, et al. “Multimodal integration learning of robot behavior using deep neural networks,” Robotics and Autonomous Systems, 62, 6, 721-736, 2014
- [3] Pin-Chu Yang, et al. “Repeatable Folding Task by Humanoid Robot Worker using Deep Learning,” IEEE Robotics and Automation Letters, 2, 2, pp397-403, 2017
- [4] M. Jonathan, et al.: “Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction”, International Conference on Artificial Neural Networks. Springer Berlin Heidelberg, 2011. p. 52-59.
- [5] Y.Yamashita, and J.Tani, “Emergence of Functional Hierarchy in a Multiple Timescales Recurrent Neural Network Model: A Humanoid Robot Experiment,” PLoS Computational Biology, 4, 11, pp.e1000220, 2008
- [6] Nextage, “Nextage-次世代産業用ロボット・ネクステージ”, 2017 年 1 月 8 日, <http://nextage.kawada.jp/>
- [7] Perception Neuron, “Perception Neuron by Noitom”, 2017 年 1 月 8 日, <https://www.neuronmocap.com/>