

## LRCN による参照点に依存した動作の認識

## Recognition of Reference-Point-Dependent Motions Using LRCN

深井 海星    武井 豪介    高瀬 健太    岩橋 直人    イエ チョウトウ    国島 文生  
 Kaisei Fukai    Gosuke Takei    Kenta Takabuchi    Naoto Iwahashi    Ye Kyaw Thu    Takeo Kunishima

岡山県立大学  
 Okayama Prefectural University

This paper evaluates the performance of Long-term Recurrent Convolutional Networks (LRCN) for the recognition of reference-point-dependent motions. The advantage of LRCN in this task is that the process of detecting landmark objects is not necessary. By experiments, we show the possibility and the problems of the use of LRCN for reference-point-dependent motion recognition.

## 1. はじめに

参照点に依存している空間的移動の概念が存在する。たとえば、図1の中で、人間によって動かされている縫いぐるみの移動は、画面中央で静止している縫いぐるみを参照点のとれば‘とびこえる’という概念であり、画面右側の箱を参照点にとれば‘のる’という概念の例になる。認知言語学 (e.g. [山梨 00]) の基本的枠組みにおいて、認知される領域において焦点化されるオブジェクトの内、相対的に際立って認知されるオブジェクトはトラジェクター、これを背景的に位置付けるオブジェクトはランドマークとして区別することで複数のオブジェクトの関係に基づく概念を記述している。図1の場合、移動している縫いぐるみがトラジェクターで、参照点として機能する静止している縫いぐるみや箱が‘とびこえる’、‘のる’の概念のそれぞれにおけるランドマークとみなせる。

こういった参照点に依存した空間的移動の概念を機械学習させる試みが古くから行われてきている。たとえば、Regier は、ランドマークに依存した動きの概念の学習について、認知科学と計算科学の二つの視点で研究を行った [Regier 96]。ここでは、ランドマークの情報が予め既知の情報として与えられるものとされていた。しかし、実際の環境ではランドマークの情報が必ずしも与えられるとは限らない。そこで、羽岡と岩橋は、ランドマークがどのオブジェクトかわからない条件下でも、参照的に依存した動作の概念を学習できる手法、参照点に依存した隠れマルコフモデル (Reference-Point-Dependent HMM (RPD-HMM)), を提案した [羽岡 00]。参照点を推定しながら動作概念を学習することを可能した。さらに、この手法の拡張として、HMM の代わりにガウス過程の隠れセミマルコフモデルを用いた手法、(Reference-Point-Dependent GP-HSMM), が提案された [Nakamura 16]。また、動作だけではなく、複数のオブジェクト間の相対的な位置関係の概念を参照点を推定しながら学習する手法の研究も行われている [田口 14]。

しかしながら、参照点を推定しながら動作や相対位置の概念を学習するこれらの手法では、事前に参照点となるオブジェクトの候補を予め用意しておくために、オブジェクト検出プロセスが必要となるのが欠点であると言える。そこで、本稿では、オブジェクト検出プロセスを必要としない参照点に依存した動作の認識法として、Long-term Recurrent Convolutional



図 1: 参照点に依存した動作の例

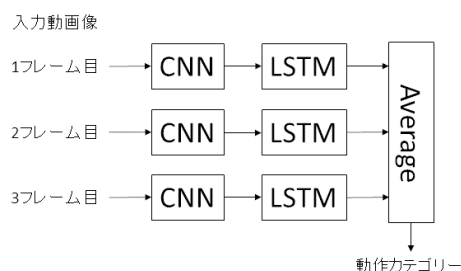


図 2: LRCN による動作概念認識プロセスの概要

Networks (LRCN) [Donahue 15] を試み、その性能を評価した。

## 2. LRCN

LRCN による動作の認識プロセスの概略は図2のとおりである。LRCN は、二種類のニューラルネットワークから構成されるディープニューラルネットワークである。1つ目は、Convolutional Neural Network (CNN) であり、入力動画画像情報を直接処理し低次元の特徴量を抽出する。2つ目は、Long Short-Term Memory (LSTM) により構成される Recurrent Neural Network であり、CNN の出力の時間構造を学習する。これら全体のネットワークのパラメータが学習データに基づいて同時に学習される。

## 3. 実験

LRCN による参照点に依存した動作の認識性能を実験により評価した。

## 3.1 データ

人間が卓上のオブジェクトを操作する様子を、RGB センサーを用いてフレームレート 30fps で収録し、実験用動画データとした。比較評価のために、参照点に依存しない動作と参照

連絡先: 岩橋直人, 岡山県立大学情報工学部, 岡山県総社市窪木 111, 0866-94-2001, iwahashi@c.oka-pu.ac.jp



図 3: “もちあげる”の例

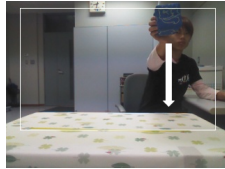


図 4: “さげる”の例

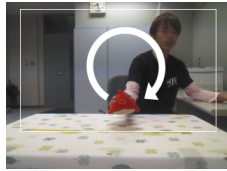


図 5: “まわす”の例



図 6: “ちかづける”の例

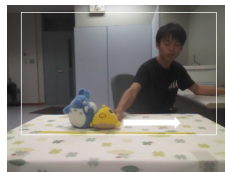


図 7: “はなす”の例

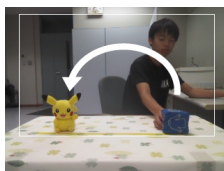


図 8: “のせる”の例



図 9: “とびこえさせる”の例

点に依存する動作を行い，以下の二つのデータベースを作成した。

**データベース A:** 参照点に依存しない動作として，“もちあげる”，“さげる”，“まわす”の3種類を，始点や終点の位置などをさまざまに変えて20回ずつ行い計60の動画像を収録した。各動画像中にあるオブジェクトの1個である。収録した動作の例を図3,4,5に示す。

**データベース B:** 参照点に依存した動作として，“ちかづける”，“はなす”，“のせる”，“とびこえさせる”の4種類を，トラジェクターとランドマークの関係をさまざまに変えて20回ずつ行い，計80の動画像を収録した。各動画像中にあるオブジェクトは2個である。収録した動作の例を図6,7,8,9に示す。

### 3.2 条件

以下の三つの条件で実験を行った。

**実験 1** データベース A 中の各動作から16個ずつの動画像，計48個をランダムに選択し，これを学習データとし残りの4個ずつの動画像，計12個をテストデータとした。

**実験 2** データベース B 中の各動作から16個ずつの動画像，計64個をランダムに選択し，これを学習データとし残りの4個ずつの動画像，計16個をテストデータとした。

**実験 3** 実験 1 と 2 の学習データを合わせた計112個の動画像を学習データとし，実験 1 と 2 のテストデータを合わせた計28個の動画像をテストデータとした。

### 3.3 結果と考察

実験結果は表1の通りである。まず，LRCN の学習が適切に行えたかを検証する。学習データ数が少ないため過学習となることを危惧していた。しかし，すべての実験における学習データとテストデータの動作認識率の差はそれほど大きくなり，大きな問題になるほどの過学習が起きなかったと言える。

次に，参照点に依存した動作の認識が適切に行えたかを検証する。実験1と実験2の動作認識率の差が小さかったことにより，参照点に依存した動作の認識が，参照点に依存しない動作の認識と同程度の性能で実現できる可能性を示している。しかしながら，実験3の動作認識率が，実験1,2に比べ，大幅に低くなっている。原因はまだ明らかではないが，訓練データの動作認識率も低いことから，学習過程においてなんらかの問題があったのではないかと考えられる。

表 1: 動作認識率 (%)

	実験 1	実験 2	実験 3
学習データ	98	97	54
テストデータ	86	81	50

## 4. まとめ

LRCN を用いた参照点に依存した動作の認識の性能を評価した。良好な性能が得られる可能性を示すとともに，本手法の問題点も明らかにした。

## 謝辞

本研究は，JSPS 科研費 15K00244，および，JST CREST 「記号創発ロボティクスによる人間機械コラボレーション基盤創成」の助成を受けて実施したものである。

## 参考文献

- [Donahue 15] Donahue, J. et al.: Long-term Recurrent Convolutional Networks for Visual Recognition and Description, in *Proc. Conf. Computer Vision and Pattern Recognition*, pp.2625-2634 (2015)
- [羽岡 00] 羽岡哲郎, 岩橋直人: 言語獲得のための参照点に依存した空間的移動の概念の学習, 電子情報通信学会技術研究報告, PRMU2000-105, pp.39-46 (2000)
- [Nakamura 16] Nakamura, T. et al.: Continuous Motion Segmentation Based on Reference-Point-Dependent GP-HSMM, in *Proc. IROS Workshop on Machine Learning Methods for High-Level Cognitive Capabilities in Robotics* (2016)
- [Regier 96] Regier, T.: *The Human Semantic Potential - Spatial Language and Constrained Connectionism*, The MIT Press (1996)
- [田口 14] 田口亮 他: SIGVerse を用いた語彙学習シミュレーション, 人工知能学会全国大会 (2014)
- [山梨 00] 山梨正明: 認知言語学原理, くろしお出版, 2000.