

1. はじめに

人間の作業工程を理解することは、ロボットを用いた作業の自動化において重要である。ロボットに作業を教示する手法として模倣学習がある。5 指ハンドを対象とした模倣は、模倣対象の行動を理解するために装着型のセンサを全ての関節に装着する。そのため、人的負担および費用的負担が大きい。本研究では、このようなセンサを必要とせず、低コストな RGB カメラから、人間の作業工程を理解する方法を提案する。

2. 提案手法

人間の作業工程を理解するためには、手検出、手の骨格検出、動作識別が要素技術として必要である。本研究では、これらを図 1 に示す 3 ステップで行う。以下に各ステップの詳細について述べる。

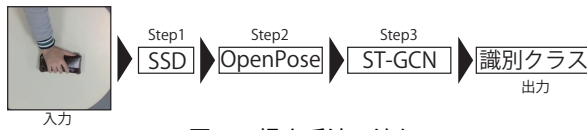


図 1：提案手法の流れ

Step1 SSD による手領域の検出

作業者の手領域を Single Shot MultiBox Detector (SSD)[1] を用いて検出する。SSD は物体検出分野において高い精度を達成しており、学習時に異なる複数のアスペクト比のバウンディングボックスを用いるため、様々な形状の物体を検出できる。検出結果を図 2 に示す。

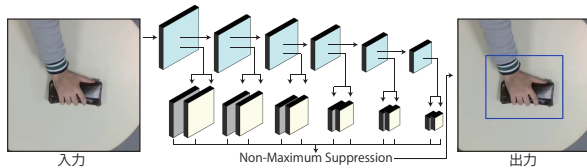


図 2：SSD による手領域検出

Step2 OpenPose による手の骨格検出

Step1 で検出した手領域を入力として、OpenPose[2] による手の骨格検出を行う。OpenPose は画像内の人間の骨格を検出する手法である。RGB 画像に対して図 3 のネットワークに示す Branch1 で関節を検出し、Branch2 で身体部位間の繋がりを表す Part Affinity Field (PAF) を推定する。検出された関節は PAF をもとに結合される。検出結果を図 3 に示す。Step1, 2 の処理は 1 フレーム単位に連続で行い、全てのフレームを対象とする。

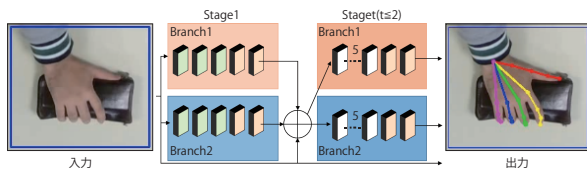


図 3：OpenPose による骨格検出

Step3 ST-GCN による手の動作識別

Step2 で得られた骨格情報を用いて Spatial Temporal Graph Convolutional Networks (ST-GCN)[3] による手の動作識別を行う。ST-GCN はノードとエッジで表現される Graph 構造の時間的な変化から、対象の動作を識別する手法である。本研究では、ノードを手の関節位置、エッジを関節の繋がりで表現し、手先の時間的な変化を捉える。ST-GCN の入力には、OpenPose で検出した手領域の手首を原点とした手関節の相対座標と信頼度を用いる。ST-GCN による畳み込みを図 4 に示す。推論結果で白く反応している部位は、ネットワークがクラス識別をする際の注視領域を示す。

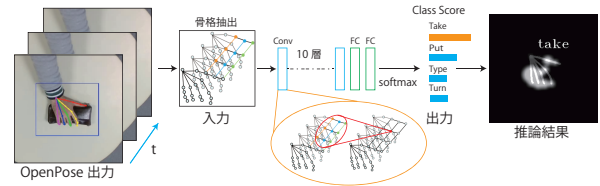


図 4：ST-GCN による畳み込み

3. 評価実験

評価実験では、Take, Put, Type, Turn の 4 クラス識別を行う。評価方法は 4 クラスそれぞれ 10 本、計 40 本の動画に対して学習モデルが正解を出した割合で評価する。Confusion matrix を用いた評価を図 5 に示す。Put の識別率が 50% と低い、これは先程述べた Take, Put, Turn の時間的骨格変化が類似していることが原因だと考える。そのため、骨格ベースの識別で類似する動作の識別率を向上させるには、追加の特徴が必要だと考える。

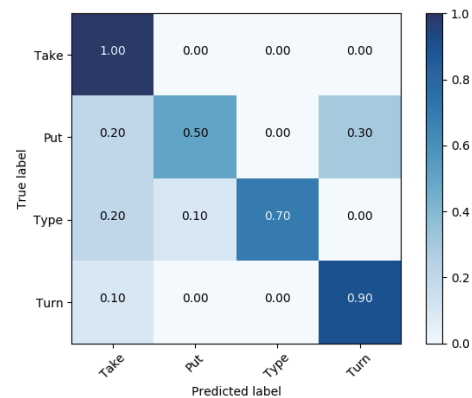


図 5：正解ラベルと推定ラベルの Confusion matrix

図 6 より Take, Put, Turn の 3 動作を識別する際に ST-GCN は、指先に注目している。これは動作時に指先以外の変化が少なかったと考えられる。対して、Take, Put, Turn と手先の形状が明らかに異なる。Type を識別する際には、各指と関節全体に注目している。

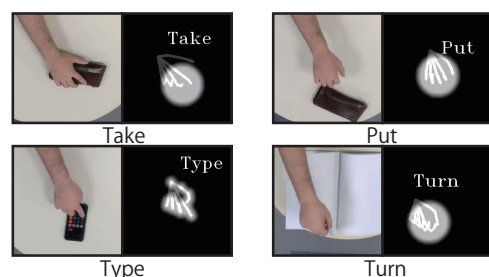


図 6：識別クラスと識別結果

4. おわりに

本研究では、5 指ハンドにおける模倣のコスト削減を目的として、1 台の RGB カメラから手検出、手骨格検出、動作識別を行う作業工程を理解する手法を提案した。今後は、Put と Turn のような動作時における骨格の時間的な変化が類似する動作に対する識別精度向上を検討する。

参考文献

- [1] W. Liu, *et al.*, “SSD: Single Shot MultiBox Detector”, ECCV, 2016.
- [2] Z. Cao, *et al.*, “Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields”, CVPR, 2017.
- [3] S. Yan, *et al.*, “Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition”, AAAI, 2018.