

デジタルヒューマンモデルを用いた 深層学習による物体組込型カメラ画像からの把持姿勢推定

稲生 健太郎[†] 家永 直人[†] 杉浦 裕太[†] 斎藤 英雄[†] 宮田 なつき^{††} 多田 充徳^{††}
[†] 慶應義塾大学 ^{††} 国立研究開発法人産業総合研究所

1. 本研究の背景

手の位置姿勢推定は近年盛んに研究がおこなわれているが、物体を持った時の姿勢を推定する研究はあまり行われていない。手は物を扱うための道具であり、実用性を考えると、その状況下で姿勢推定が行われるべきである。本稿では、カメラ画像からの手の物体把持姿勢推定について検討する。

単視点 RGB 画像より手の把持姿勢推定を行ったものとして、柏木らの手法[1]がある。この手法では物体上に組み込まれた魚眼カメラに映る指先位置と手の中心の3次元位置を求め、参照用の姿勢モデルを1つ用意し、推定される位置に合わせるという方法を取っている。この手法の問題点は、指が5本触れていないと計測ができない、ということである。物体の持ち方、例えば野球ボールであれば薬指や小指を曲げて持つ方法もあれば、小指だけ曲げる持ち方も考えられ持ち方に制限がある。本手法では、ボールを様々な形で握った様子を多数集めてデータセットとし、機械学習させることで持ち方に関する制約を無くすことを目標とする。

2. 提案手法

直径 10cm の球を様々な姿勢で握り、モーションキャプチャで指先や関節の位置の情報を記録した。これを Dhaibaworks[2]とよばれる、モーションキャプチャの情報に合うような手の CG モデルを出力するソフトウェアを用い、図1のように仮想空間に再現した。ボール上の手と反対側にカメラが付けられていると仮定して、図2のようなカメラ画像を 5810 枚作成した。その際、各画像に対して指先や関節直上といった手のキーポイント計 28 個の3次元位置をモーションキャプチャの情報から取得し、各学習データにアノテートしたものを CNN で学習した。



図1. 手モデルの出力例

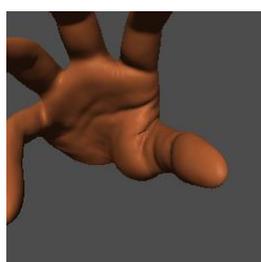


図2 学習データ

学習済みの CNN により、手が映った入力画像から手のキーポイントの3次元位置を推定する。推定した位置をプロットし、指ごとに直線でつなげることで大まかな姿勢を求める。

3. 評価実験

学習データと同じ状況で取得し、学習には用いなかったテストデータ 115 枚を入力として検証した結果、キーポイント位置の推定誤差平均は約 12mm であった。関連手法[3]では誤差平均が約 19mm であり、最新の研究と同等またはそれ以上の結果であった。

また、一般的なカメラで母指、示指、中指で球を握った手を撮影した画像(図 3)を入力とし、姿勢を推定した結果を図4に示す。図中の赤、青、黄緑、黒、黄の線はそれぞれ母指、示指、中指、環指、小指を表す。環指と小指を曲げ、母指を横に、示指を上、中指を斜めに伸ばすように球を持った様子を推定できている。



図3 入力画像

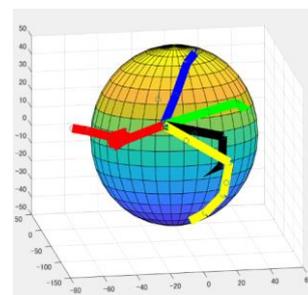


図4 推定結果

4. 今後の課題

データセット、学習ネットワークの調整を行うことで推定精度の向上を図る予定である。

参考文献

- [1] 柏木ら “組込み型カメラによる把持動作時の人体手形状 3D 復元.” 映像情報メディア学会誌, vol. 71, no. 11, pp. J283-J286, 2017
- [2] DhaibaWorks 公式サイト, <http://www.dhaibaworks.com/> (accessed 2018-02-07)
- [3] Christian Zimmermann and Thomas Brox, “Learning to Estimate 3D Hand Pose from Single RGB Images”, IEEE International Conference on Computer Vision (ICCV), 2017