



2020年10月21日

東京工業大学
科学技術振興機構

手首装着型小型カメラによるジェスチャー認識技術を開発

ースマホ操作や手話認識などへの応用を期待ー

【要点】

- 手首装着型小型カメラ1台のみで実現可能な、手指ジェスチャー認識技術の開発に成功。
- 深層学習で3次元手指姿勢を推定し、高い認識精度を実現。
- 小型カメラ搭載スマートウォッチを使用した、スマホ操作や手話認識などへの応用を期待。

【概要】

東京工業大学 情報理工学院 情報工学系の小池英樹教授の研究グループは、手首装着型小型カメラで撮影した手の甲のカラー画像を**深層学習ネットワーク**（用語1）で学習することで、手指のジェスチャーを認識する技術を開発した。

従来手指認識は、固定したカメラで撮影するか、手袋型デバイスを使用する必要があり、前者は手指の稼働範囲が制限され、後者は手袋着用の手間があった。本成果によって、認識範囲の制限がなくなり、例えば、手指のジェスチャーだけでスマートフォンのアプリケーションを起動する、音量を変える、といった操作ができるようになる。具体的な使用デバイスとしては、小型カメラ搭載のスマートウォッチなどを想定している。

今回の成果は、東京工業大学の他、米国カーネギーメロン大学ロボティクスインスティテュートの **Kris M. Kitani** 教授、オーストラリア New South Wales 大学の **Aaron Quigley** 教授らとの共同研究によるもので、国際会議「The 33rd Annual ACM Symposium on User Interface Software and Technology (UIST 2020)」で10月23日（現地時間）に発表される。

●背景

手指のジェスチャー認識は、家電機器の操作、コンピュータや仮想現実感 (VR) における操作入力、あるいは手話認識など広い分野への応用が期待されている。従来技術として、1 台以上の固定されたカメラを利用したジェスチャー認識はあるが、カメラの前の限られた範囲内ではしか認識ができず、例えば歩行中のジェスチャーなど、広い範囲での認識はできなかった。一方、手袋型デバイスを装着する手法もあるが、手袋着用の手間があること、デバイスが高価であること、一般的な使用には向かないなど問題点がある。

●研究成果

本研究で開発した手指ジェスチャー認識システムでは、利用者の手首の甲側に小型カメラを装着する (図 1)。具体的な使用デバイスとしては、竜頭の部分に小型カメラを搭載したスマートウォッチを想定している。この小型カメラで撮影した映像には、利用者の手の甲のみが映っている。この画像を入力として、撮影時の手指の 3 次元姿勢を出力とする深層学習ネットワークを新たに設計・開発した。

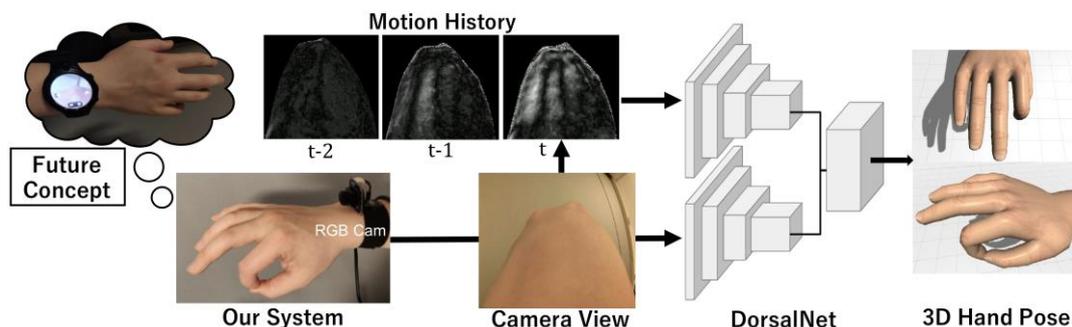


図 1. システムの概略。手首の甲側につけた小型カメラの映像だけで手指の 3 次元姿勢を推定する。

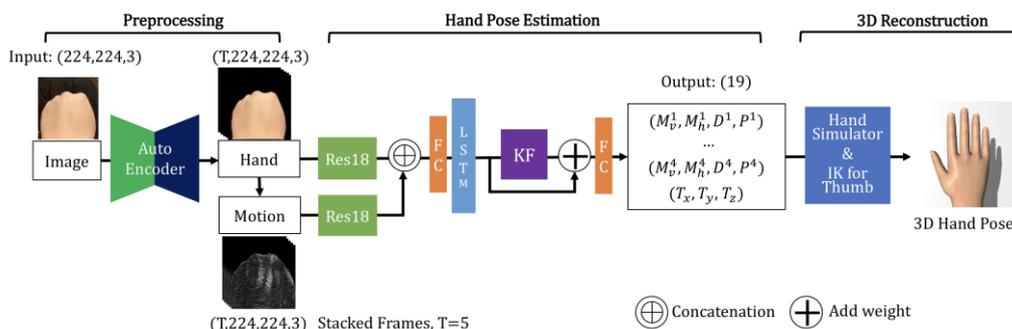


図 2. 3 次元手指姿勢認識用深層学習ネットワーク

カメラが撮影した手の甲のカラー画像と、そのカラー画像を2値化処理した白黒画像の過去数フレームを、それぞれ深層学習ネットワーク **ResNet18** (用語 2) に入力し、両方の出力を**全結合層 (FC)** (用語 3)、**長短期記憶層 (LSTM)** (用語 4)、**カルマンフィルタ層 (KF)** (用語 5)、全結合層 (FC) で処理することにより、指関節の3次元座標が出力される。これを手指の3次元姿勢再構成モデルで処理することで、3次元手指姿勢を推定する。

評価実験の結果、静的なジェスチャー認識で最高 91.4%、タッピングなどの動的なジェスチャー認識で最高 89.4%の認識精度を達成した (表 1)。今後、データセットの拡充でさらなる精度向上が期待できる。

表 1. 静的ジェスチャーと動的ジェスチャーの認識精度の評価実験結果

Methods	Static ASL Gesture Accuracy		
	Individ.	General	Leave-1-user
Nearest N. [13]	44.6%	37.4%	33.4%
Direct Regression	71.2%	69.0%	57.2%
Yeo et al. [53]	88.6%	82.5%	70.4%
Ours (w/o KF)	91.4%	88.8%	84.8%
Ours (w/ KF)	90.2%	88.8%	83.0%

Methods	Dynamic Tapping Accuracy		
	Individ.	General	Leave-1-user
Nearest N. [13]	47.8%	40.0%	28.4%
Direct Regression	51.0%	50.4%	39.8%
Yeo et al. [53]	59.6%	62.2%	55.0%
Ours (w/o KF)	85.2%	85.0%	77.6%
Ours (w/ KF)	89.4%	86.8%	79.8%

●今後の展開

小型カメラを搭載したスマートウォッチはすでに市販されている。今後は、同様に小型カメラを搭載した腕輪型デバイスも普及すると考えられ、それに伴って、今回開発された手指のジェスチャー認識技術が広く利用されると期待される。

【用語説明】

- (1) 深層学習ネットワーク：人間の神経細胞の仕組みをモデルとしたニューラルネットワークを多層化した機械学習システム。
- (2) **ResNet18**：ResNet は Microsoft Research が開発した一般物体認識用ニューラルネットで、18 は階層の数を表す。
- (3) 全結合層 (FC)：入力層のノードと出力層のノードが全て結合されているネットワーク。

- (4) 長短期記憶層 (LSTM) : 時系列データの処理に用いられるネットワーク。
- (5) カルマンフィルタ層 (KF) : 動的システムの予測に用いるネットワーク。

【発表予定】

- ・国際会議 : Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology (UIST2020)
<https://uist.acm.org/uist2020/>
- ・タイトル : Back-Hand-Pose: 3D Hand Pose Estimation for a Wrist-worn Camera via Dorsum Deformation Network
- ・著者 : Erwin Wu, Hui-Shyong Yeo, Aaron Quigley, Hideki Koike, Kris Kitani

【付記】

- 今回の研究成果は、以下の事業・研究領域・研究課題によって得られた。
- ・科学技術振興機構 (JST) 戦略的創造研究推進事業 チーム型研究 (CREST)
 - ・研究領域 : 人間と情報環境の共生インタラクション基盤技術の創出と展開
(研究総括 : 間瀬 健二 (名古屋大学 大学院 情報学研究科 教授))
 - ・研究課題名 : 技能獲得メカニズムの原理解明および獲得支援システムへの展開
 - ・研究代表者 : 小池 英樹 (東京工業大学 情報理工学院 教授)
 - ・研究期間 : 平成 29 年 10 月 ~ 令和 5 年 3 月

【問い合わせ先】

東京工業大学 情報理工学院 情報工学系 教授
小池英樹

Email: [koike\[at\]c.titech.ac.jp](mailto:koike@c.titech.ac.jp)

TEL: 03-5734-3526 FAX: 03-5734-3526

(JST 事業に関すること)

科学技術振興機構 戦略研究推進部 ICT グループ
舘澤博子

E-mail: [crest\[at\]jst.go.jp](mailto:crest[at]jst.go.jp)

TEL: 03-3512-3526 FAX: 03-3222-2064

【取材申し込み先】

東京工業大学 総務部 広報課

E-mail: [media\[at\]jim.titech.ac.jp](mailto:media[at]jim.titech.ac.jp)

TEL: 03-5734-2975 FAX: 03-5734-3661

科学技術振興機構 広報課

E-mail: [jstkoho\[at\]jst.go.jp](mailto:jstkoho[at]jst.go.jp)

TEL: 03-5214-8404 FAX: 03-5214-8432