



Tokyo Tech



2020 年 10 月 21 日

報道機関各位

東京工業大学
科学技術振興機構

胸装着型小型カメラ 1 台によるモーションキャプチャ技術を開発

ー広い運動範囲でアニメーション制作などへの応用を期待ー

【要点】

- 胸に装着した小型超魚眼カメラ 1 台のみで撮影可能なモーションキャプチャ技術の開発に成功。
- 深層学習により利用者の身体形状を推定、同時に一人称視点映像の合成も可能。
- 運動範囲の広いモーションキャプチャシステムとして、スポーツ科学、医療、アニメーション制作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らの研究グループは、胸装着型の小型超魚眼カメラ（用語 1）で撮影した映像を深層学習ネットワーク（用語 2）で学習することで、利用者の身体形状を正確に推定するモーションキャプチャ技術を開発した。さらに、身体形状と同時に頭部の向きを推定し、一人称視点映像を合成することを可能にした。

従来の光学式モーションキャプチャでは、複数台の高性能カメラを部屋の天井や壁に固定し、利用者はマーカー付きスーツを着用して計測する必要があるため、利用者の運動範囲は狭い空間内に制限された。また、一人称視点映像の合成には頭部にカメラを別途装着する必要があった。今回開発した手法では、運動範囲の制限がないため、スポーツ科学、医療、アニメーション制作などの分野でのモーションキャプチャの利便性が高まると期待される。さらに、使用機材が小型カメラ 1 台のみであるため、コストダウンも可能である。

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 21 日（現地時間）に発表される。

●背景

モーションキャプチャ技術は、スポーツ科学、医療、アニメーション制作などで広く用いられており、現在は光学式と慣性式が主流である。光学式モーションキャプチャシステムでは、同期させた複数台の高性能カメラを部屋の天井や壁に固定して、動作計測を行う。しかしこのシステムは、運動範囲がカメラで撮影する空間内に限定されるため、利用者が広い空間を移動する場合には適用できなかった。また、利用者はマーカの付いたスーツを着用しなければならず、利便性も低かった。一方、慣性式モーションキャプチャシステムでは、各関節部分に慣性センサを付けたスーツを着用し、動作計測を行う。運動範囲は広いが、周辺に金属があると測定誤差が生じる。また、スーツ着用の手間がかかる点も光学式システムと変わらない。さらに、光学式、慣性式ともに、導入コストは数百万円以上と高額である。

近年、利用者に小型カメラを装着し、撮影した映像を深層学習ネットワークに学習させることで、利用者の動作推定をする研究が行われている。しかし、これまで開発された手法では、利用者の頭部にカメラを装着するために装着感がよくないという問題があった。またカメラの画角が小さいため、上肢部分のみの推定に限定されていた。

●研究成果

本研究で開発したモーションキャプチャシステムでは、利用者の胸に画角 280 度の超魚眼カメラを装着する（図 1）。この超魚眼カメラで撮影した映像には、身体形状の推定に必要な、利用者の頭部、両手、両足が全て映っている。この超魚眼映像を入力とし、撮影時の 3 次元姿勢と一人称視点を出力とする深層学習ネットワークを新たに設計した。

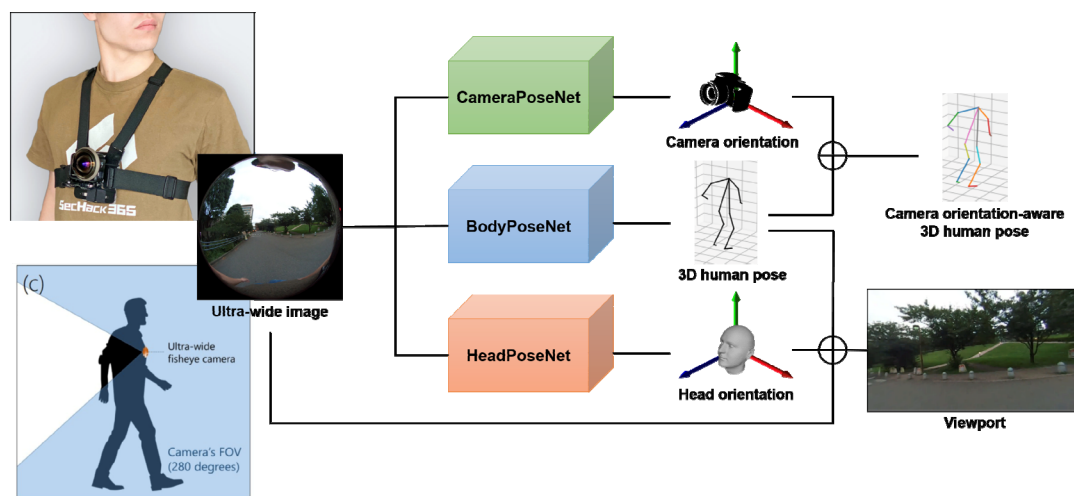


図 1. システムの概略。胸につけた超魚眼カメラ映像だけで 3 次元姿勢を推定し、かつ視点映像を合成する。

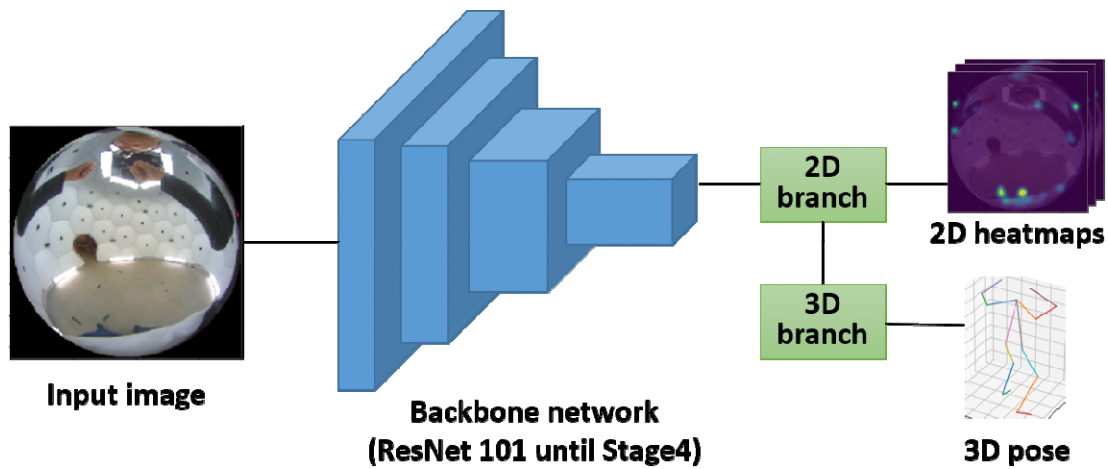


図 2. 3 次元姿勢推定用 BodyPoseNet のアーキテクチャ

深層学習のためには、人工データと実データのデータセットが必要になる。本研究では、3 次元 CG を用いて作成した 680,000 枚の超魚眼映像と 3 次元姿勢からなる人工データと、実際に撮影した 16,000 枚の超魚眼映像と 3 次元姿勢からなる実データを用意した（インターネット上で公開予定）。

評価実験の結果、人工データでの各関節における平均誤差は、上肢部分で 24.6mm、下肢部分で 62.5 mm、全体として 43.6 mm となった（表 1）。一方、実データでは平均誤差が 84.9 mm となったが、これは今後のデータセットの拡充で精度向上が期待できると考えられる。

表 1. 関節の推定精度の評価実験結果. 人工データ(MonoEye dataset)では平均誤差が 43.6 mm

Category	MonoEye dataset	Real-world dataset								
	All	Walking	Sitting	Crawling	Crouching	Boxing	Waiving	Stretching	Kicking	All
Upper body	24.6	75.3	69.7	108.6	69.6	86.6	130.7	133.7	77.6	94
Lower body	62.5	71.1	71.3	88.6	66.2	75.1	77.8	83.5	72.0	75.7
Average	43.6	73.2	70.5	98.6	67.9	80.9	104.3	108.6	74.8	84.9

また、本システムでは利用者の頭部姿勢を推定し、一人称視点映像を合成することも可能にした。推定された頭部姿勢の情報を用いて、超魚眼映像内から利用者が見ている部分を検出し、一人称視点映像として合成する（図 3）。

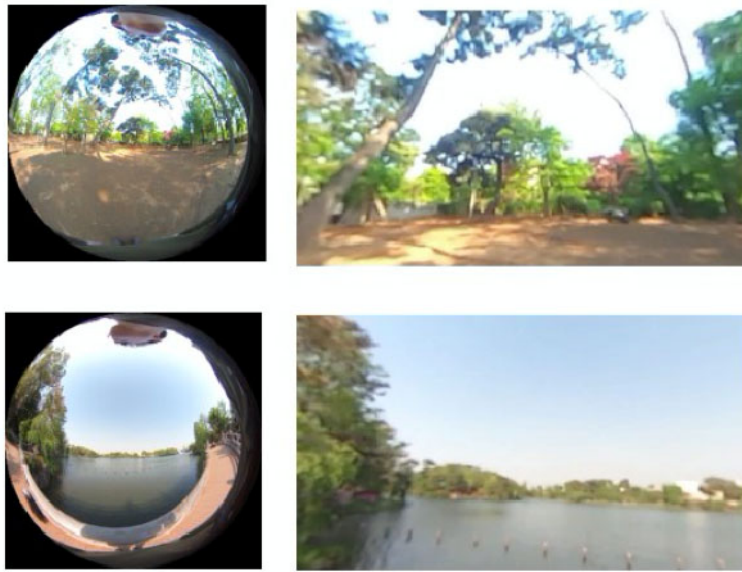


図 3. 超魚眼映像と頭部姿勢から合成された一人称視点映像

●今後の展開

本研究で開発したモーションキャプチャシステムは、カメラの小型化によって利便性が向上するとともに、コストダウンも可能であることから、スポーツ科学、医療、アニメーション制作など幅広い分野での利用が期待される。今後はデータセットの拡充により、計測精度のさらなる向上を目指す予定である。

【用語説明】

- (1) **超魚眼カメラ**：超魚眼レンズを搭載したカメラ。一般的な魚眼レンズは画角が 180 度程度であるが、使用した超魚眼レンズは画角 280 度である。
- (2) **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。

【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: MonoEye: Multimodal Human Motion Capture System Using A Single Ultra-Wide Fisheye Camera
- ・著者: Dong-Hyun Hwang, Kohei Aso, Ye Yuan, Kris Kitani, Hideki Koike

【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

E-mail: 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

館澤博子

E-mail: crest@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@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@jst.go.jp

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