

戦略的創造研究推進事業 AIP 加速課題
研究課題「バイオ実験作業の知の獲得と
ロボットへの転写」

研究終了報告書

研究期間 2020年4月～2023年3月

研究代表者：佐藤 洋一
(東京大学生産技術研究所、教授)

§1 研究実施の概要

(1) 実施概要

少子高齢化社会先進国である日本において、高齢化が進む国民の健康長寿を担うという意味でバイオ産業は最も重要な産業分野の一つであるが、そのバイオ産業の研究開発の現場において、高い技能を持つバイオ実験技術者の不足という課題が顕在化している。これは、バイオ実験では繊細かつ単純な作業が多く労働集約的である上に、高度な技能の獲得に経験と勘が必要で人材育成が容易ではないということが一因となっている。さらに、近年、いわゆる実験の「再現性の危機」が広く認識されるようになった。この問題には、実験結果が経験と勘を含めた個々の実験技術者の力量に依存してしまうことに加えて、論文中の実験プロトコルでは非言語的情報が記述できず、実験の再現に必要な情報が決定的に不足していることが深く関与している。このようなことから、我が国におけるバイオ産業の健全な発展のためには、バイオ実験の大幅な省力化と再現性の問題の解決が急務となっている。

このような背景の下、本研究課題では、バイオ実験技術者の人材育成ならびにバイオ実験の省力化と再現性向上に資するべく、バイオ実験作業の技能に関する知の獲得とロボットへの転移の実現を目指し、1)バイオ実験作業記録映像データセットの構築、2)バイオ実験作業認識技術の開発、3)バイオ実験技術者の技能のモデリング技術の開発、4)バイオ実験技術者からバイオ実験ロボットへの作業技能転写の4つの研究項目に取り組んだ。

研究項目1. バイオ実験作業映像データセット構築

- バイオ実験作業オントロジー構築(産総研 Gr)
- バイオ実験作業映像の収集(産総研 Gr、東大 Gr)
- バイオ実験作業映像のアノテーション(産総研 Gr、東大 Gr)

研究項目2. 実験作業認識技術の開発

- 固定視点と一人称視点映像の統合による作業認識(東大 Gr)
- 異なる一人称視点映像の統合による作業認識(CMU Gr)

研究項目3. 実験作業技能モデリング技術の開発

- スキルレベル定量評価・可視化(東大 Gr)
- 実験器具3次元トラッキング(CMU Gr)

研究項目4. バイオ実験ロボットへの作業技能転写(東大 Gr、CMU Gr、産総研 Gr)

その結果、得られた主要な成果は以下の通りである。バイオ実験作業映像データセット構築に関して、32名の実験技術者が7種の実験プロトコルを行う作業の様子を、頭部装着型カメラと5台の固定カメラで撮影した実験作業映像を収集し、プロトコル、タスク、動作、物体の異なる階層で密なアノテーション付与を実施した。このデータセットは、収録人数、プロトコル数、実験時間、延べ映像時間ともに最大であり、実験作業を一人称視点映像と複数外部視点映像で同時記録している点においても他に類を見ない。また、実験作業認識技術の開発では、特に、動作認識と動作セグメンテーション(検出と識別)の頑健化と高精度化、少量データからの動作認識、手作業解析における手領域抽出の頑健化、手と物体のコンタクト状態検出、被操作物体検出の高精度化に関して主要な成果を得た。さらに、実験作業技能モデリング技術の開発では、作業スキルレベルの定量的推定と可視化、動作認識モデルの振る舞いの可視化、複雑シーンにおける頑健かつ高速な物体3次元トラッキングについて成果を得ることができた。また、バイオ実験ロボットへの作業技能転写に関して、人のデモンストレーションに基づくロボットの動作モデルの学習で主要な成果を得ている。本研究課題で得られたこれらの成果は、トップ国際会議やジャーナルに多数採択され、その新規性と有効性が高く評価されている。今後は、バイオ実験作業に限らず、さまざまな応用シーンにおいて、映像からの人物行動理解とロボット動作学習の要素技術として広く活用されることが期待される。

(2) 顕著な成果

<優れた基礎研究としての成果>

1. 前後動作の文脈に基づく高精度な動作セグメンテーション

概要: 一人称視点映像からの重要動作区間の検出・認識において、カメラ姿勢が大きく変動する場合に動作を上手く捉えられないという課題が存在した。これに対し、本研究では、前後動作の関係性のモデリングにより動作検出・認識をリファインするという独自の着想に基づき、**Graph-based Temporal Reasoning Module (GTRM)**と呼ばれる手法を提案し、既存手法を大幅に越える精度での動作検出・認識を達成した。また、この手法は様々な既存手法と組み合わせ利用可能という特長も有する。この成果はトップ会議 CVPR 2020 に採択され多数引用されている。

2. 環境変動に頑健な一人称視点映像からの手領域抽出

概要: 一人称視点映像からの詳細動作解析のための手領域抽出という基本タスクに関して、**Bayesian CNN** に基づく手領域検出モデルの不確かさを手掛かりとした教師無しドメイン適応手法を提案し、異なる環境下で撮影された一人称視点映像に対して新たに教師用データを用いることなく、既存手法を越える手領域検出精度を達成した。また、複数モデルによる予測のコンセンサスに基づく疑似ラベル生成という発想に基づく手法を開発し、さらなる性能向上を実現した。これらの成果はトップ会議 CVPR 2020、ECCV 2022、国際ジャーナル IEEE Access に採択され、特許出願も行った。

3. モデル進化の枠組みに基づく人のデモンストレーションからのロボット動作学習

概要: 人が行うデモンストレーションに基づくロボット動作学習では、人の手とロボットハンドの間の形状や関節の位置・数・動作範囲などの違いにどのように対処するかが根本的な課題となる。これに対し、手からロボットハンドへのモデルの進化という考え方に基づき、強化学習によりハンドモデルの進化とハンド制御に関するポリシーの獲得を行うことにより、人によるデモンストレーションから任意のロボットハンドへの動作の転写を実現した。これは人からロボットへの作業スキル転写の実現に向けて、人の手とロボットハンドのギャップという困難な課題の解決につながる重要な成果であり、機械学習とロボティクス分野のトップ会議である ICML 2022 と CoRL 2022 に採択された。

<科学技術イノベーションに大きく寄与する成果>

1. 物体 3 次元姿勢推定データセット StereOBJ-1M の構築

概要: ステレオ画像からの物体 3 次元姿勢推定のための大規模データセット StereOBJ-1M を構築し公開した。StereOBJ-1M は、既存の物体 3 次元位置・姿勢推定データセットの多くが距離画像を対象としているのに対し、ステレオ RGB 画像を扱う唯一のデータセットであり、画像フレーム数とアノテーション数に関して最大規模となっている。さらに、透明物体や鏡面反射状の物体を扱っているという点や、屋外シーンを含む 180 地点以上でのデータを含んでいるという点においても他に類を見ない。StereOBJ-1M は、このコンピュータビジョンにおける重要タスクである物体 3 次元姿勢推定に関する研究に広く資することが期待される。この成果はトップ会議 ICCV 2021 に採択された。

2. バイオ実験作業映像データセットの構築

概要: 2 つの異なるバイオ実験室環境において、実験技術者が行う実験作業の様子を頭部装着型カメラと周囲に複数配置した固定カメラで捉えたバイオ実験作業映像データセットを構築した。2 つの実験室環境において 32 名の実験参加者が 7 つのプロトコルを実施する様子を記録した映像に対し、プロトコル、タスク、動作、物体の異なる階層で密なアノテーションが付与されている。データセットの規模は、一人称視点と外部視点の映像セット 226 組、総実験作

業時間約 13 時間、延べ映像長さ約 78 時間となり、収録人数、プロトコル数、実験時間、延べ映像時間ともに最大となっている。また、実験作業を一人称視点映像と複数外部視点映像で同時記録している点も他に類を見ず、本プロジェクトに限らず広く研究コミュニティに資する研究資源として活用されることが期待される。

3. セマンティック統合による動作スキル推定

概要: 熟練技術者から非熟練技術者への作業技能の伝承を図るには、作業における動作スキルのモデリングと定量的評価が求められる。本研究では、両手でツールを用いて行う作業の動作スキルレベル推定について、映像から意味のある要素を自動的に発見し、それらの時系列的な関係性をモデリングという着想に基づく手法を開発し、外科手術訓練を対象としたスキルレベル評価ベンチマークデータセットを用いた評価において、既存手法を大幅に越える精度を達成した。この成果はその新規性と有効性が高く評価され、トップ会議 MICCAI 2022 で採択された。映像からの動作スキル解析という挑戦的課題の解決に向けて、重要な一歩となることが期待される。

< 代表的な論文 >

1. Yifei Huang, Yusuke Sugano, and Yoichi Sato, “Improving Action Segmentation via Graph Based Temporal Reasoning,” in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2020), pp. 14024-14034, June 2020.

概要: 映像からの行動理解に関する基本タスクの一つであるアクションセグメンテーションについて、Graph Convolutional Networks により動作間の関係性を考慮するという独自の着想に基づき、高精度な動作区間検出および動作識別を実現したものであり、代表的な一人称視点映像ベンチマークデータセットを用いた実験において、既存手法を大幅に越える性能を達成した。この成果はトップ会議 CVPR 2020 に採択され多数引用されている。また、この手法は様々なベースライン手法と組み合わせることで利用可能という特長を有しており、一人称視点映像からの動作検出・認識タスクにおいて広く活用されることが期待される。

2. Takehiko Ohkawa, Yu-Jhe Li, Qichen Fu, Ryosuke Furuta, Kris Kitani, and Yoichi Sato, “Domain Adaptive Hand Keypoint and Pixel Localization in the Wild,” in Proc. European Conference on Computer Vision (ECCV 2022), pp. 1-20, October 2022.

概要: 手動作解析において重要な手領域抽出と手指姿勢推定のタスクについて、我々が先に開発した Consensus Pseudo Labeling の枠組みを教師-生徒ネットワークモデルによる自己訓練の考え方でさらに発展させることにより、新たな環境下での学習用教師データを全く必要とせず、複数のベンチマークデータセットで既存手法を越える高精度な手領域抽出と手指姿勢推定を実現したものである。この成果は提案手法の新規性と有効性が高く評価されトップ会議 ECCV 2022 に採択された。

3. Xingyu Liu, Deepak Pathak, and Kris M. Kitani, “HERD: Continuous Human-to-Robot Evolution for Learning from Human Demonstration,” in Proc. Conference on Robot Learning, pp. 1-12, December 2022.

概要: 人によるデモンストレーションにもとづくロボット動作学習において、人の手とロボットハンドの形状や自由度の違いが課題となる。これに対し、本論文では人の手指のモデルからロボットハンドのモデルへの進化という考え方に基づき、強化学習の枠組みでハンドモデルの進化とハンド制御に関するポリシーを効率的に獲得することを他に先駆けて実現した。これは人からロボットへの作業スキル転写の実現に向けて、人の手とロボットハンドのギャップという困難な課題の解決につながる重要な成果であり、ロボットと機械学習に関するトップ会議 CoRL 2022 に採択された。

§ 2 研究実施体制

(1) 研究チームの体制について

① 東大グループ

研究代表者: 佐藤 洋一 (東京大学生産技術研究所 教授)

研究項目

- バイオ実験記録映像データセット構築
 - バイオ実験記録映像の収集
 - バイオ実験記録映像のアノテーション
- バイオ実験作業認識技術の開発
 - 固定視点と一人称視点の統合による作業認識
- バイオ実験作業技能モデリング技術の開発
 - スキルレベル定量評価・可視化
- バイオ実験作業映像の構造化とバイオ実験ロボットへの作業技能転写

② 産総研グループ

主たる共同研究者: 光山 統泰 (産業技術総合研究所人工知能研究センター 研究チーム長)

研究項目

- バイオ実験記録映像データセット構築
 - バイオ実験作業オントロジーの構築
 - バイオ実験記録映像の収集
 - バイオ実験記録映像のアノテーション
- バイオ実験作業映像の構造化とバイオ実験ロボットへの作業技能転写

③ CMU グループ

主たる共同研究者: Kris M. Kitani (The Robotics Institute, School of Computer Science, Carnegie Mellon University, Associate Research Professor)

研究項目

- バイオ実験作業認識技術の開発
 - 異なる一人称視点映像の統合による作業認識
- バイオ実験作業技能モデリング技術の開発
 - 実験器具 3 次元トラッキング
- バイオ実験作業映像の構造化とバイオ実験ロボットへの作業技能転写

(2) 国内外の研究者や産業界等との連携によるネットワーク形成の状況について

本研究プロジェクトが基盤とする一人称視点映像解析技術に関連して、東大グループと CMU グループが大規模一人称視点映像データセット構築プロジェクト Ego4D (<https://ego4d-data.org>) に参画し活動を進めてきている。このプロジェクトは、世界各国約 15 の有力研究機関により構成される国際コンソーシアムにより実施されており、本研究プロジェクトで取り組んだ研究開発の内容とも密接に関係したものとなっている。Ego4D プロジェクトに関連した協働として、コンソーシアムメンバーによる論文の共同執筆、Ego4D データセットを用いたベンチマークによるコンテストの実施、Ego4D データセットに関する国際ワークショップの実施などがある。また、本研究プロジェクトからの情報発信とネットワーク形成に寄与する取り組みとして、CMU グループの Kris M. Kitani が国際会議 MVA 2021 において、映像からの人物行動理解に関するチュートリアルを実施した。