

先端国際共同研究推進事業
2023 年度採択
「Top 研究者/Top チームのための ASPIRE」
AI・情報分野

2023 年度～2024 年度
年次報告書（公開版）

研究課題名	思いやり AI 創生に向けた人の基盤モデル構築のための 国際プラットフォーム
日本側研究代表者	西野 恒 京都大学 教授
相手側研究代表者	・Vincent Lepetit, Professor, École de Ponts ・Angjoo Kanazawa, Assistant Professor, University of California Berkeley ・Jianbo Shi, Professor, University of Pennsylvania ・Todd Zickler, Professor, Harvard University
研究期間	2024 年 2 月 1 日～2029 年 3 月 31 日

1. 研究成果の概要

① 研究構想にかかる成果

＜実施したこと＞

本年度の研究計画の中でも大きな実施項目として、本研究計画の全ての研究チームで構築し共有する人間の行動観察データのコアとなるデータ取得を行なった。6 日間に渡り計 50 人（男性 27 人、女性 23 人）の 20～60 代の被験者が自然な室内環境で自然な日常動作を行う様子を撮影した。撮影は国際高等研究所にある家族用住宅内で行った。撮影環境は、リビング-ダイニング・廊下からなり、リビング-ダイニングには多数の家具を配置した。各日 6～8 人の実験者により撮影準備・被験者の案内を行い、被験者 1 人につき 1 回 1～17 分（平均 4.8 分）の撮影を 3～9 回（平均 3.5 回）行なった。撮影全体には計 48 時間要し、最終的に 14.4 時間分のデータセットとなった。このデータセットは、部屋の周囲を取り囲む 27 台の時間同期されたカメラで撮影された 129 種の動作を含む動画と、各フレームの人体姿勢と手の姿勢のラベル、人体形状のラベルから構成される。これらのデータ撮像には日本側チーム研究員、博士課程学生、並びに修士課程学生全員が参加したばかりなく、UC バークレイの博士課程学生の招聘に合わせて実施することにより、彼らの参加によるデータ取得ノウハウの共有も行なった。

また、新たに画像からの物体 3 次元形状とカメラ姿勢の同時推定手法を仏 ENPC 研究チームと共同で導出した。車のように模様がなく、視点依存な光沢を持つ既存手法では取り扱えない物体でも、較正等の必要なく形状復元を行うべく、物体表面における光の鏡面反射自体を利用することにより、反射によって映り込んだ世界の中で対応点群を検出することにより、カメラ姿勢および物体形状を一意に推定する手法を確立した。これは、本研究プログラムで目指す部屋環境における人物行動理解のために、特に対象人物と物体の相互作用を正確に理解するためには欠かせない基盤技術となりうる。この研究成果は、その新規性が高く評価され、コンピュータビジョン分野における難関会議である European Conference on Computer Vision において数少ない（全採択論文中 2.3%）オーラル発表に採択され、イタリア・ミラノにおいて発表を行った。さらに、相手側ハーバード大学研究チームと、新たな物体形状復元の手法を導出した。本研究プログラムの目指す客観視からの人間の行動理解を実現する上において、環境に存在する人間が操作しうる物体の 3 次元形状を簡便かつ正確に推定する技術が必須となる。そこで、新たに単一画像から物体の精緻な形状を、その多義性も含め復元する方法を導出した。この成果は機械学習のトップ会議である Neural Information Processing Systems (NeurIPS)において、特に注目に値する研究である Spotlight（全論文中 8.1%）として採録され、バンクーバーにおいて学会発表を行なった。現在、さらに人間視覚の多義性メカニズムの解明に向け、本成果を応用すべく共同研究を継続している。

関連する形状復元手法として、本研究プログラムで考慮する室内環境の復元においても重要な基盤技術となる単眼深度推定を視点に依存せずに自動的にスケールを較正する手法を導出した。この研究成果はコンピュータビジョンの難関会議である European Conference on Computer Vision (ECCV)において発表を行い、筆頭著者である博士課程学生の初めての海外研究発表となった。

日本側研究チームは、さらに物体の見えから周囲環境の見えおよび物体反射特性の同時推定、ならびに偏光を用いたパターン光の投影による物体形状及び反射特性復元をおこなう新たな手法をそれぞれ導出した。前者は確率的拡散過程を用いた定式化により、反射特性によりぼやけた見えから復元しうる周囲環境をサン

プリングにより網羅的に推定することを実現し、後者に関してはアクティブセンシングであるものの、目には見えず人に気づかれない推定を実現した。これらの成果はコンピュータビジョンのトップ会議である IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)において学会発表を行い、特に前者は特に注目すべき論文に与えられる Highlight 発表の機会を得た。これらの成果に基づき、さらにハーバード大学研究チームと両者の深化による、日常環境への適用を計画している。

このように、コンピュータビジョンおよび機械学習のトップ会議等において、相手側チームとの共著論文 2 件を含む計 6 本の研究論文が採択され、これらの研究成果を通して本研究プログラムの基盤技術の高い水準での確立およびトップ研究者たちへの周知が世界的に行えたばかりではなく、ASPIRE プログラム自体も広く知らしめることができた。

② 国際ネットワーク構築・拡大／国際頭脳循環の促進に資する若手研究者の人材育成に関する成果

日本側チームの特定研究員が 2 回に分け延べ 3 カ月以上フランス ENPC の相手側研究チーム IMAGINE に滞在し、研究活動を行った。IMAGINE では、コンピュータビジョンとその周辺分野において世界を牽引する研究が行われており、滞在中は共同研究者である ENPC 教授 Vincent Lepetit と週 1 回程度打ち合わせを行い、共同研究を進めた。1 回目の滞在中は、自然な光源環境下において生じる鏡面反射を利用した、カメラ姿勢と物体形状の同時推定手法の評価・改良に取り組み、2 回目の滞在中は、物体の動きによる視点変化を手掛かりとした、疎な多視点動画像からの 3 次元形状復元に関する定式化について取り組んだ。週 1 回の研究室セミナーや 3 泊 4 日の合宿形式のワークショップ「IMAGINE Retreat」、および研究室に所属する研究者や学生との日常的な交流を通して、欧州における研究活動への理解および研究内容の相互理解を深めた。

相手側研究チームの PI 及び学生の多くが京都大学に滞在し、日本側研究チームと多くの交流及び研究遂行した。フランス ENPC 研究チームからは博士課程学生 2 名が 1 カ月程度滞在し、それぞれ 3 次元復元および物体姿勢推定の研究を日本側チームと共同して推進した。これらの研究は、当該学生たちの帰国後も週一回のミーティングを通して継続しており、現在論文投稿に向けた準備を行なっている。米国 UC バークレイ研究チームからは博士課程学生 4 名が 1 カ月滞在し、主に非剛体物体の映像から形状および素材推定に関する研究を日本側チームと立ち上げ進めた。現在も週一回のリモートミーティングを通して日本側チームの研究員および修士学生との研究を継続しており、論文投稿へ向けた準備を行なっている。さらに、UC バークレイ研究チームから博士研究員が 1 カ月滞在し、日本側研究チーム修士課程学生と連携し人間の衣服を含めた動的 3 次元モデル化の研究に従事した。この研究も現在週 1 回のリモートミーティングをもとに進められており、論文投稿準備に取り組んでいる。また、ペンシルバニア大学 Jianbo Shi 教授の 2 回にわたる延べ 2 カ月の滞在により、自己視点映像を用いた研究に関する深い議論と共同研究基盤の構築を行い、現在修士課程学生を中心に研究を進めている。

さらに、日本側研究チームの博士課程学生 3 名、修士課程学生 6 名を帯同し、コンピュータビジョン分野トップ会議である IEEE/CVF Conference on Computer Vision and Pattern Recognition に参加し、世界トップレベルの研究に関する見聞および学生たちとの交流を含めた。これらの経験や相手側チーム学生の招聘を通じた日々密接に連携しつつ直接共同研究を行うことにより、日本側チームの学生が触発され、博士課程進学を決定したり、検討している学生 3 名やすでに博士課程進学を行なった学生 2 名が生まれるなど、進路決定のための大きな材料となっている。また、国内研究会（情報処理学会コンピュータとイメージメディア）に発表

のない学部 4 回生を 5 名参加させ、同様に、博士課程進学検討の動機づけが得られた。

以上のように、日本側チームから研究員の延べ 7 カ月程度の海外滞在、相手側研究チーム学生や PI 延べ 8 名の延べ 9 カ月以上に渡る招聘滞在により、日本側チーム全ての学生が常に関連する研究において海外トップ研究者たちとの深い交流を実現でき、博士進学を決意するなど、将来の頭脳循環につながる実りの多い影響を与えることができています。

③ 国際頭脳循環の促進に資する若手研究者の人材育成に関する成果

本研究プログラムは単一の日本側チームで構成される統合された研究テーマに基づき研究を遂行しているため、①研究構想にかかる実施内容・成果に詳述した通りの研究テーマを実施した。すでに詳細を記述した通り、年次計画書の計画通り全ての研究テーマを実施できた。

- 1) データ取得およびモデル化プラットフォームためのデータ取得を実施し、日本側チーム総動員で取り組んだばかりでなく、米国側チームの実地参加により具体的ノウハウの共有を実現した。
- 2) また、精緻かつ駆動可能な形状モデルの導出に関し、特に物体形状の復元に関する質の高い研究（コンピュータビジョントップ会議および機械学習トップ会議で発表済み）を相手側チームと共同で進めたばかりではなく、計画通り駆動可能な人体形状モデルの基礎導出を実施した。
- 3) さらに、意図発言の視覚理解基盤として、相手側チームと共同研究により、自己視点映像を用いた新たな解析基盤技術を現在構築中である。上述のデータ取得による大規模データのアノテーション等を順次進めており、これらを待って実世界人間理解への活用を引き続き行なっていく。

2. 研究実施体制

研究テーマ	中心となる研究者氏名	所属機関・部署・役職名
人の基盤モデル構築	西野 恒 Angjoo Kanazawa (UC Berkeleyからは他にJitendra Malik, Alexei A. Efros, Trevor Darrellも参加)	京都大学、情報学研究科、教授 UC Berkeley, EECS, Assistant Professor
	Vincent Lepetit Jianbo Shi Todd Zickler	École des Ponts, CNRS, Professor Univ. of Pennsylvania, CIS, Professor Harvard University, EECS, Professor

3. 代表的な業績（原著論文、プレスリリース、表彰など）

原著論文

- Kohei Yamashita, Vincent Lepetit, and Ko Nishino, Correspondences of the Third Kind: Camera Pose Estimation from Object Reflection, European Conference on Computer Vision ECCV’24, Oct., 2024.

- Xinran Nicle Han, Todd Zickler, and Ko Nishino, Multistable Shape from Shading Emerges from Patch Diffusion, Advances in Neural Information Processing Systems NeurIPS'24, 2024.
- Yuto Enyo and Ko Nishino, Diffusion Reflectance Map: Single-Image Stochastic Inverse Rendering of Illumination and Reflectance, Conference on Computer Vision and Pattern Recognition CVPR'24, Jun., 2024.
- Yuto Enyo and Ko Nishino, Diffusion Reflectance Map: Single-Image Stochastic Inverse Rendering of Illumination and Reflectance, Conference on Computer Vision and Pattern Recognition CVPR'24, Jun., 2024.
- Genki Kinoshita and Ko Nishino, Camera Height Doesn't Change: Unsupervised Training for Metric Monocular Road-Scene Depth Estimation, European Conference on Computer Vision ECCV'24, Oct., 2024.