様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　５ 年　 ９ 月 ２０ 日現在

機関番号：８２６２６
研究種目：基盤研究(B)（一般）
研究期間：2020～2022
課題番号：２０Ｈ０４２１７

研究課題名（和文）4Dアースキャプショニング

研究課題名（英文）4D Earth Captioning

研究代表者
櫻田　健（Sakurada, Ken）

国立研究開発法人産業技術総合研究所・情報・人間工学領域・主任研究員

研究者番号：７０７７３６７０

交付決定額（研究期間全体）：（直接経費）　　13,700,000円

研究成果の概要（和文）：本研究では、動画像等から街並みの３次元地図の構築および更新を効率的に行い、地上で起こっている出来事を人間が理解しやすいテキストで説明するための基盤技術を開発した。この技術により、将来、自動運転や運転支援のために搭載したカメラ、ドライブレコーダー、スマートフォンなどから収集した膨大なデータを、３次元地図やテキスト等の形でユーザーに提示し活用できる可能性を示した。さらに、マップ共有時に重要な課題となるプライバシー保護に対して、シーンプライバシーに配慮したリアルタイムの３次元地図構築技術を開発した。

研究成果の学術的意義や社会的意義
コンピュータビジョンやロボティクス分野では、街並みの変化を検出する研究は行われてきたが、多くの場合３次元地図の更新を目的としているため、画素や点群、ボクセルのような観測単位の変化のみに着目しており、その意味的な理解はほとんど注目されてこなかった。本研究では、シーンを物体の集合として仮定し、物体単位の変化の検出と説明文生成を行うことで、人間が理解しやすく応用タスクで扱いやすい形で提示することを可能とした。また、シーンプライバシー保護手法において、従来の点群マップと異なる直線群マップに対し新たな再投影誤差モデルを定義し、カメラ姿勢と点群・直線群マップを同時に最適化可能とした点も学術的意義が大きい。

研究成果の概要（英文）：We have developed fundamental technologies for efficiently reconstructing and updating 3D maps of a city from images and for explaining events occurring on the ground in a text that is easy for humans to understand. These technologies have shown the possibility of utilizing the vast amount of data collected from smartphones, drive recorders, and cameras installed for autonomous driving and driver assistance, where these data will be presented to users in the form of 3D maps and text. In addition, we have developed a real-time 3D localization and mapping method that considers scene privacy, which is a concern in map sharing.

研究分野：コンピュータビジョン

キーワード：画像説明文生成　変化検出　空間モデリング

１．研究開始当初の背景

地球上で起こっている現象を捉えるために、コンピュータビジョンやリモートセンシングの分野では、航空・衛星などの画像から地物の判定を行う研究が古くから行われてきた。近年では、車載カメラなどの画像を計測の範囲や分解能について相補的に利用し、地上の状態をより詳細に計測する方法も開発されている。これら地上の様子を解析する研究の背景には、自動運転用地図の作成やインフラの点検、災害対応、農業の自動化など大きな社会的ニーズがあり、複数の種類、時刻、スケールの計測データを統一的に扱うためのデータベースとその枠組みが次第に整備されつつある。

そのデータベースから空間的な表現だけでなく、時間的な変化も考慮してシーン（風景）を記述するのが 4D モデリングである。過去数十年の写真から街並みの 2D あるいは 3D 的な変化の可視化、衛星や車載のカメラ画像を用いた地震や津波の被害推定など、多くの研究が活発に行われている。一方で、広域の現象を迅速に理解するためには、面積や体積、特定物体の有無を数値で記述するだけでなく、膨大なデータから検出した現象がどのようなものであるかを人間が分かりやすい形で整理・提示する必要がある。

２．研究の目的

本研究では、運転用地図の作成やインフラの点検、災害対応、農業の自動化など大きな社会的ニーズがあり、複数の種類、時刻、スケールの計測データから、地球上で起きている現象を自動で検出し、任意視点の説明文を生成する「4D アースキャプショニング」を目指した。文章で記述することにより、大量データの把握が容易となり、さらに、関係性も含めたテキストによる検索も可能となる。

街並み（の変化）の認識はこれまでにも取り組みがなされてきた。これまでは撮影視点に対して認識精度が大きく依存したが、本研究により、車載カメラのように機械的に撮影し毎回撮影視点が異なる画像などに対してもロバストにシーンを文章として記述することができる。また、都市あるいは地方スケールのモデリングをするためには、多時刻かつ広域の画像を地理的に正しく登録することが必須である。

３．研究の方法

シーン変化の検出や説明文生成を学習・評価するためには、人や車などの移動物体以外のシーン変化を含む異なる時刻の画像ペアを大量に取得する必要がある。そのため、本研究では主に自動運転用シミュレーターである CARLA Simulator[1]を利用して、大量のシーン変化を含む画像ペアと変化ラベルを生成した。

さらに、シーン変化の説明文生成を学習・評価するためには、シーン変化の説明文を大量に作成する必要がある。そこで本研究では、Amazon Mechanical Turk の作業者に対して、図１のように変化の仕方、相対的な位置関係、カテゴリに関して英語で説明文を作成するように依頼し、67000 個以上のキャプションを取得した。本データセットを CARLA-Change-Captioning データセットと呼び、training ： validation ： test の３セットに分割し、各手法の定量評価に利用した。

図１．シーン変化の説明文を付与するための作業画面

また、シーンプライバシーの保護機能を有する３次元復元手法の開発においても、オーバーラップを有するカメラ軌跡の動画が必要となるため、同様に CARLA Simulator[1]を利用してデータセットを作成した。このシミュレーション動画でカメラ姿勢の定量評価を、さらに、実環境を撮影した動画データで定性評価を行い、提案手法の有効性と頑健性を確認した。

[1] A. Dosovitskiy et al.,“CARLA: An Open Urban Driving Simulator”, The 1st Annual Conference on Robot Learning, 2017

４．研究成果

本研究の主な成果として、(１)物体レベルのシーン変化検出、(２)シーン変化の説明文生成、(３)プライバシー保護機能を有する Visual SLAM について以下で説明する。

(１) 物体レベルのシーンの変化検出

画像を用いてシーン変化を検出する場合、既存手法の多くでは、画素単位の変化を推定していたため、物体単位の意味的な変化を検出する場合、独立に学習した物体検出器と組み合わせた後処理が必要である上に、画素単位の対応付けは照明条件やカメラ視点の変化に脆弱であった。特に、車載カメラのような移動体に搭載されたカメラの場合、撮影毎に視点が異なるため画素間の対応付けや遮蔽関係の考慮が難しい問題があった。

この問題を解決するため、本研究では、物体検出器から抽出した物体の特徴量をグラフマッチング問題の枠組みで対応づけることでシーン変化をロバストに検出する手法を開発した（図２）。提案手法は、事前学習された物体検出モジュールとグラフニューラルネットワークを統合することで、物体単位の変化検出を End-to-end で学習することを可能とした。図３は提案手法による推定結果を示しており、一般的な２次元地図や自動運転用の３次元地図を更新する上で重要な構成要素である建物や道路標識、ポールなどの変化を、一組の画像ペアから検出することが可能となった。

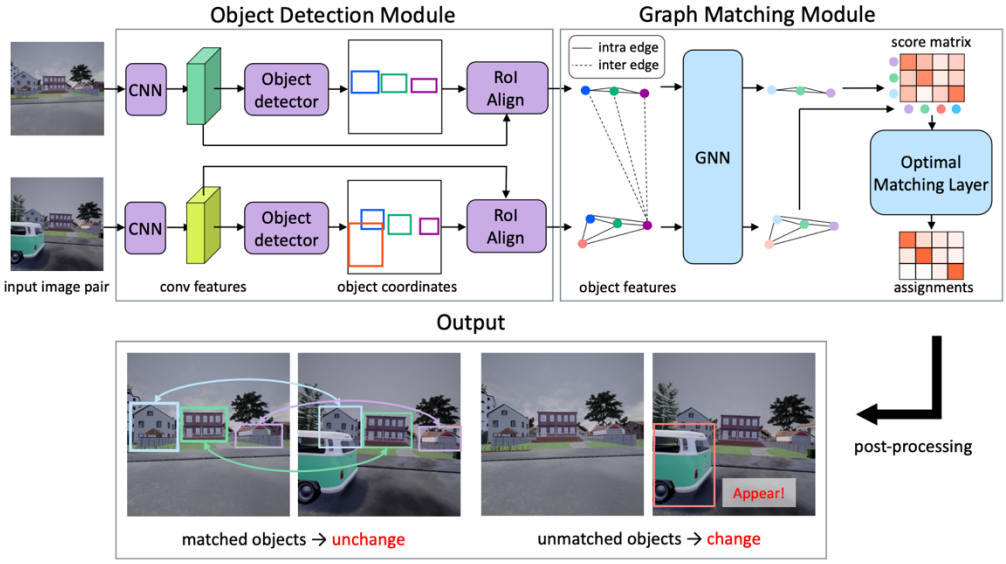


図２．物体レベルの変化検出ネットワークの概要

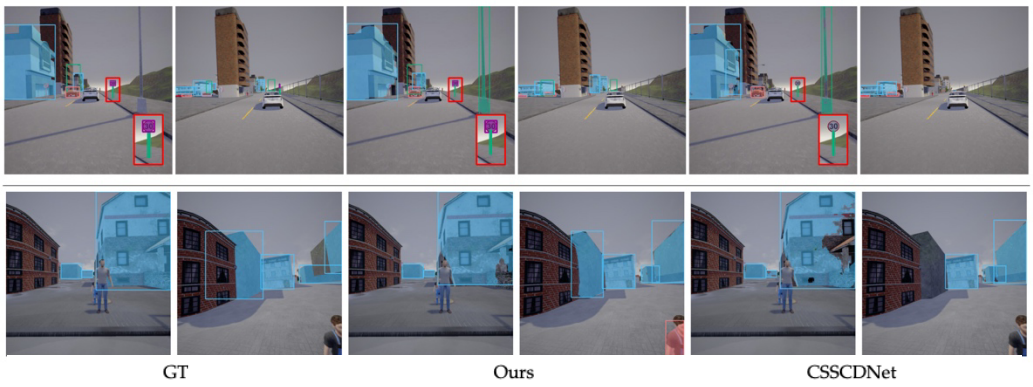


図３．物体レベルの変化検出結果

(２) シーン変化の説明文生成

シーン変化を人間が理解しやすいテキストで出力する先行研究はいくつか存在するが、その多くは画像をグリッドで表現し、シーンを構成する物体単位の情報を十分に活用できていなかった。 そのため、ストリートシーンのような前景背景の分離が難しいより複雑な環境では、物体の変化を正しく捉えた説明文を生成することが困難であった。

この課題を解決するため、図４のような、物体間の関係性を抽出して記述するネットワークを構築した。本研究で新たに提案した Object Relational Image Encoder (ORIE)は、物体検出器で抽出した物体の特徴量を transformer encoder を用いて、画像内外の関係性を考慮した特徴計算を行うネットワークであり、その特徴量から画像全体の、あるいは各物体に関する変化の説明文を生成することができる。

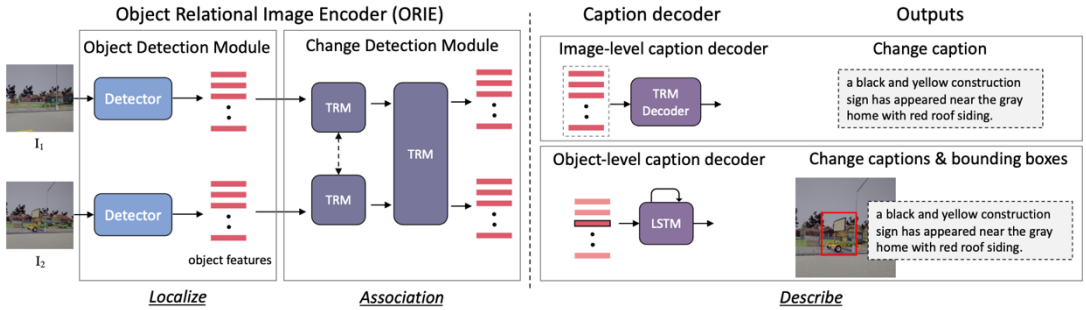


図４．物体の特徴量を利用した画像説明文生成ネットワーク

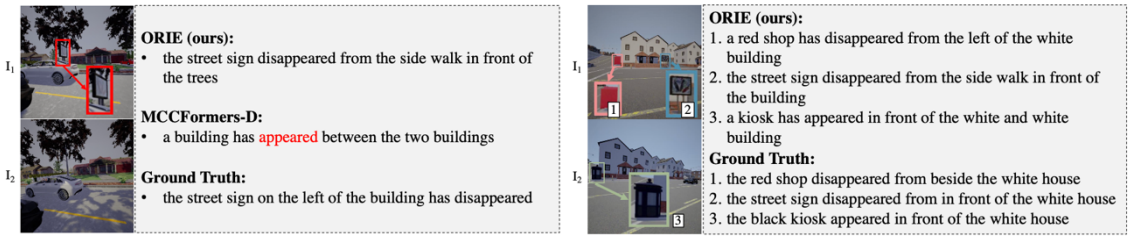


図５．CARLA-Change-Captioning データセットおける画像レベル（左）と物体レベル（右）のシーン変化説明文生成の結果

図５に CARLA-Change-Captioning データセットにおける、画像レベルと物体レベルのシーン変化説明文生成の結果を示す。画像レベル（左）の場合、提案手法では Ground Truth と同様に、建物の左に写っている道路標識が無くなったことを正しく記述できている。また、物体レベルの変化（右）においても、店や道路標識の変化をそれぞれ正しく説明できていることが分かる。本提案手法により、車載カメラ画像のようなより複雑なシーンに対して、変化の説明文を自動生成するための基盤技術を確立した。

（３）シーンプライバシーの保護機能を有する Visual SLAM

自動運転やロボティクス、xR などにおいて、シーンの３次元地図を異なるユーザーやエージェント間で共有し、大域的な自己位置を推定することは重要な要素技術の一つである。動画象を用いた３次元復元では、画像から検出した特徴点を多視点間で対応付け、疎な３次元特徴点群としてシーンを復元する方法が最も実用的に用いられている。しかし、疎な３次元点群を画像平面に投影し Convolutional Neural Network (CNN)へ入力することで、元のシーン画像を復元（反転攻撃）できることが明らかにされている[2]。

このシーンプライバシーの問題を解決する方法とし、３次元点群をランダムな方向の直線群に置き換えることで、位置推定機能を保ちつつも反転攻撃を難しくする手法が提案されている[3]。しかし、３次元点をランダムな方向の直線に置き換えることは１次元分の情報を欠落させることであり、画像のカメラ姿勢を推定するための拘束条件も３次元点群マップの場合より多く必要となるため、計算量の増加と精度の低下を引き起こし、動画等のリアルタイムアプリケーションへの適用は困難となる。

そこで、本研究では、計算効率を考慮したリローカリゼーションと 2D-3D マッチング手法を提案し、事前に与えられた３次元直線群に対するリアルタイムのトラッキングおよびマッピングを可能とした（図６）。さらに、３次元直線に対する新たな再投影誤差（図７）を定義し、３次元直線と３次元点が混在するマップに適用可能な各種バンドル調整手法を新たに提案した（図８）。その一つである Global BA では、３次元直線が画像上の観測を有していない問題に対し、擬似観測を定義することで、サーバーから与えられた３次元直線群を含む全体最適化を可能とした（図９）。

[2] F. Pittaluga et al., “Revealing Scenes by Inverting Structure From Motion Reconstructions”, CVPR, 2019
[3] P. Speciale et al., “Privacy preserving image-based localization”, CVPR, 2019

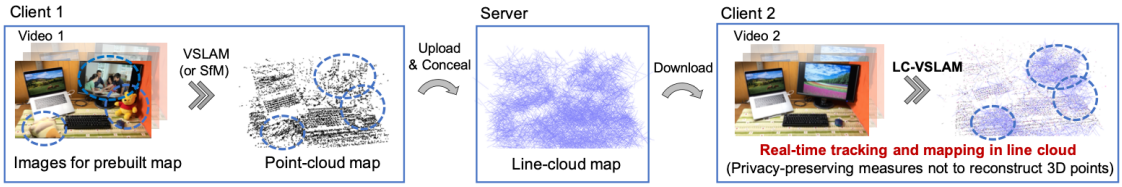


図６．シーンプライバシーの保護機能を有する Visual SLAM の概要

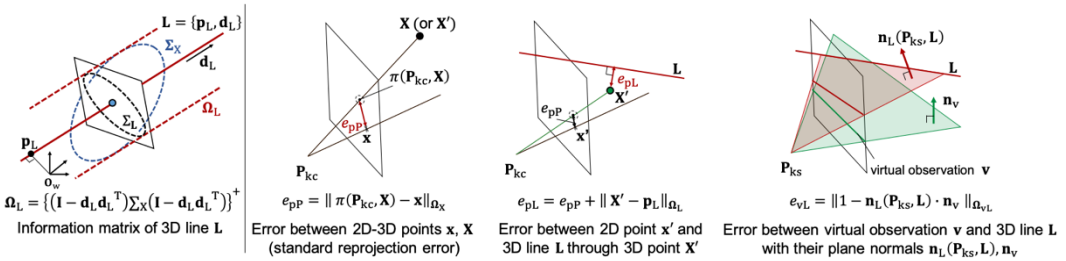
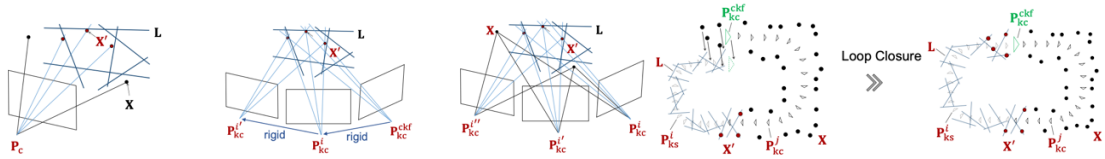
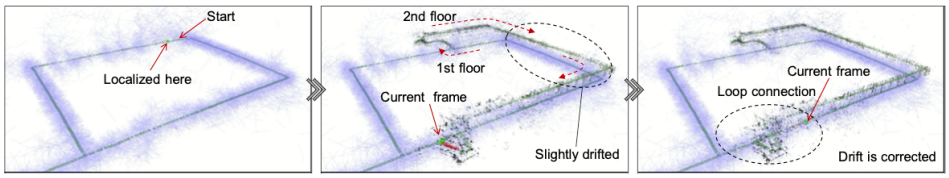


図７．３次元直線に対する新たな再投影誤差



(a) Motion-only BA (b) Rigid-stereo BA (c) Local BA (d) Global BA before and after loop closure

図８．３次元直線と３次元点が混在するマップに適用可能な各種バンドル調整手法



(a) Relocalization (b) Before loop closure (c) After loop closure

図９．サーバーから与えられた３次元直線群も含む全体最適化の様子

５．主な発表論文等

〔雑誌論文〕　計5件（うち査読付論文　5件／うち国際共著　1件／うちオープンアクセス　4件）

１．著者名 Doi Kento、Hamaguchi Ryuhei、Iwasawa Yusuke、Onishi Masaki、Matsuo Yutaka、Sakurada Ken	４．巻 14
２．論文標題 Detecting Object-Level Scene Changes in Images with Viewpoint Differences Using Graph Matching	５．発行年 2022年
３．雑誌名 Remote Sensing	６．最初と最後の頁 1～20
掲載論文のDOI（デジタルオブジェクト識別子） 10.3390/rs14174225	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 Hamaguchi Ryuhei、Furukawa Yasutaka、Onishi Masaki、Sakurada Ken	４．巻 -
２．論文標題 Heterogeneous Grid Convolution for Adaptive, Efficient, and Controllable Computation	５．発行年 2021年
３．雑誌名 Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)	６．最初と最後の頁 13946 - 13955
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/CVPR46437.2021.01373	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

１．著者名 Shibuya Mikiya、Sumikura Shinya、Sakurada Ken	４．巻 -
２．論文標題 Privacy Preserving Visual SLAM	５．発行年 2020年
３．雑誌名 Proceedings of European Conference on Computer Vision	６．最初と最後の頁 102～118
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-3-030-58542-6_7	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 Furukawa Yukuko、Suzuki Kumiko、Hamaguchi Ryuhei、Onishi Masaki、Sakurada Ken	４．巻 -
２．論文標題 Self-supervised Simultaneous Alignment and Change Detection	５．発行年 2020年
３．雑誌名 Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems	６．最初と最後の頁 6025～6031
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/IROS45743.2020.9340840	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

1．著者名 Doi Kento、Sakurada Ken、Onishi Masaki、Iwasaki Akira	4．巻 -
2．論文標題 GAN-Based SAR-to-Optical Image Translation with Region Information	5．発行年 2020年
3．雑誌名 Proceedings of IEEE International Geoscience and Remote Sensing Symposium	6．最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/IGARSS39084.2020.9323085	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

〔学会発表〕　計9件（うち招待講演　0件／うち国際学会　0件）

1．発表者名 土居 健人、濱口 竜平、岩澤 有祐、大西 正輝、松尾 豊、櫻田 健
2．発表標題 Pixel vs. Object: 変化キャプショニングにおける最適な画像表現についての研究
3．学会等名 第25回 画像の認識・理解シンポジウム
4．発表年 2022年

1．発表者名 土居健人，岩澤有祐，松尾豊，大西正輝，櫻田健
2．発表標題 CARLAシミュレータを用いた変化キャプショニング既存手法のベンチマーク
3．学会等名 第24回画像の認識・理解シンポジウム(MIRU)
4．発表年 2021年

1．発表者名 市原光将，渋谷樹弥，大川快，大西正輝，櫻田健
2．発表標題 バイナリ超平面を利用した高速な次元削減手法の提案
3．学会等名 第24回画像の認識・理解シンポジウム(MIRU)
4．発表年 2021年

１．発表者名
渋谷 樹弥，角倉 慎弥，横田 理央，櫻田 健

２．発表標題
プライバシー保護を考慮したVisual SLAM

３．学会等名
第23回画像の認識・理解シンポジウム(MIRU)

４．発表年
2020年

１．発表者名
濱口 竜平，古川 泰隆，大西 正輝，櫻田 健

２．発表標題
Graph Residual Networks for Semantic Segmentation

３．学会等名
第23回画像の認識・理解シンポジウム(MIRU)

４．発表年
2020年

１．発表者名
金子 真也，櫻田 健，池畑 諭，相澤 清晴

２．発表標題
三次元構造エッジ推定に基づく深層学習を用いた単眼画像からのシーンメッシュ復元

３．学会等名
第23回画像の認識・理解シンポジウム(MIRU)

４．発表年
2020年

１．発表者名
古川 悠久子，鈴木久美子，濱口 竜平，大西 正輝，櫻田 健

２．発表標題
自己教師あり学習による位置合わせとシーン変化の同時推定

３．学会等名
第23回画像の認識・理解シンポジウム(MIRU)

４．発表年
2020年

１．発表者名
芝 慎太朗, 金子 真也, 青木 義満, 櫻田 健

２．発表標題
Unsupervised dense depth estimation from event camera

３．学会等名
第23回画像の認識・理解シンポジウム(MIRU)

４．発表年
2020年

１．発表者名
岩瀬 駿, 横田 理央, 櫻田 健

２．発表標題
エピポーラ拘束付きの深層グラフマッチングを用いた視点変化に堅牢な変化検出手法の提案

３．学会等名
第23回画像の認識・理解シンポジウム(MIRU)

４．発表年
2020年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

Privacy Preserving Visual SLAM
https://xdspacelab.github.io/lcvslam/

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
研究分担者	高村　大也 (Takamura Hiroya) (80361773)	国立研究開発法人産業技術総合研究所・情報・人間工学領域・研究チーム長 (82626)	

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関			
カナダ	Simon Fraser University			