

科学研究費助成事業 研究成果報告書

令和 5 年 6 月 5 日現在

機関番号：13301

研究種目：若手研究(B)

研究期間：2017～2022

課題番号：17K12712

研究課題名（和文）線虫の神経細胞に対する細胞核自動アノテーションのための人工知能技術の開発

研究課題名（英文）Automatic Annotation of Neural Cells in C.elegans Imaging Data

研究代表者

広瀬 修 (Hirose, Osamu)

金沢大学・生命理工学系・准教授

研究者番号：30549671

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：本研究で、線虫の神経細胞核撮影データに対する細胞核アノテーション自動化技術の開発を行った。線虫は原始的な脳を有しており、その全神経細胞レベルでの活性の定量化は脳情報処理メカニズムの解析のステップとして重要である。実施者らは、線虫の全神経細胞規模の撮影データから神経細胞活性の自動定量化手法をすでに開発していた。一方で、撮影された神経細胞核が解剖図上のどの細胞なのかを決定すること（アノテーション）は、神経細胞1つ1つの役割を理解する上で不可欠であるものの、非常に難しく専門家の労力を要する課題であった。本研究でこれらの自動化技術の開発を目指した。関連成果をまとめた論文2報が一流誌に掲載された。

研究成果の学術的意義や社会的意義

本研究で、線虫の細胞核イメージングデータに対して撮像された神経細胞それぞれの名前を自動特定する手法の開発を行った。開発手法は当時の最高精度のアノテーション手法の精度には残念ながら及ばなかったものの、この手法を改善させることを目的として行った研究が大きな成果を挙げた。開発手法の核は「形状位置合わせ」を行うアルゴリズムであり、研究期間の後半期はこれらのアルゴリズムの改善に注力した。結果として、形状解析やコンピュータグラフィックスにも応用可能である汎用アルゴリズムの開発につながった。開発したソフトウェアを実施者のホームページで配布しており、現在も世界中からダウンロードされている。

研究成果の概要（英文）：To understand information processing in an organism's brain, we quantify neural activities using specialized imaging technologies for C. elegans. Our automated techniques enable the segmentation and tracking of neuronal nuclei, allowing us to quantify neural activities throughout the entire brain. An important task is neuron annotation, which involves identifying neurons from imaging data. Neuron annotation is challenging due to varying neuron positions caused by body posture differences, even in nematodes with the same posture. Our goal is to automate neuron annotation using C. elegans imaging data. The data types we analyze include 3D positions of neurons obtained through automatic detection and manual annotations by experts. We address neuron annotation as a point set matching problem, proposing three algorithms: Hungarian, coherent point drift, and our novel dependent landmark drift. We evaluate algorithm performance using partially annotated location data of C. elegans neurons.

研究分野：統計科学

キーワード：画像解析 細胞画像 細胞名特定 形状解析 位置合わせ 点群マッチング

1. 研究開始当初の背景

(1) 生きた線虫の神経細胞活性の自動定量化技術が実施者らの貢献により確立された

実施者らは生きた線虫の神経細胞核のみを視覚化した 3 次元動画像データ(図 1 左; 九州大学石原健教授の提供)から神経細胞核を検出・追跡する手法を開発した(図 1 中・右). これは線虫の脳情報処理解析に向けて各々の神経細胞の活性を定量化することが目的で, 線虫全体の動きに応じて細胞核も移動するため, それらを追跡する必要があるからである. 実施者はこのデータに対する細胞核の自動追跡手法に関する論文を報告し, 開発したソフトウェアを実施者のウェブサイト <http://github.com/ohirose/spf/> にて配布している.

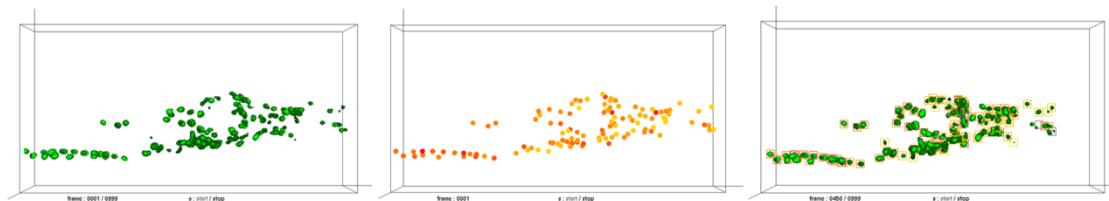


図 1: (左)線虫の神経細胞核だけを経時的に 3 次元撮影したデータ. 楕円体形状物体が 1 つの神経細胞核である. (中)3 次元静止画像に対する細胞核の検出・数え上げ. 比較的低輝度の核も数え上げに成功している. (右)3 次元動画像 400 フレーム目に対する追跡結果. 400 フレームを経過してもほとんどの細胞核の追跡に成功している.

(2) 神経細胞核アノテーションの自動化が線虫の大規模な脳情報処理研究に必要不可欠である

実施者らの研究により, 生きた線虫の神経細胞活性の自動定量化が可能になったが, 線虫の脳情報処理研究を行う上で重要な課題がある: 自動検出した細胞核が神経細胞配置図 (図 2) 上のどの細胞に対応するのかが分からないことである. 線虫の神経細胞は刺激への応答性に個性があり, 刺激応答性の違いや存在する場所によりその 1 つ 1 つに名前が付けられている. ここでは撮影データ中の神経細胞核に神経細胞配置図上の名前を対応づける作業をアノテーション(細胞名特定)と呼ぶ. 神経細胞核のアノテーションは,

- (1) 撮影時の線虫の体位による細胞位置の違い,
- (2) 細胞が密集しその境界が不明瞭な画像領域の存在,
- (3) 細胞位置の個体差,
- (4) 公開されている神経細胞の位置情報の不正確さのため, 線虫研究の専門家であっても易しくはない作業である.

また, 頭部の神経細胞は 200 個程度あり撮影データは 3 次元画像で 1000 枚規模であるため, 専門家の手によるアノテーションの作業負荷は非常に高い. 一方で, 環境からの刺激に対する応答を全神経細胞レベルで解析するには相当数回の全神経細胞アノテーションが必須である. そのため, 線虫の専門家の知識・経験を代替する人工知能技術の開発が急務であった.

2. 研究の目的

線虫神経細胞の精密な次元配置図の作成とアノテーション自動化のための人工知能技術の開発

本研究で, 今後の線虫の脳情報処理研究の土台作りを目的として以下を行った.

1. 部分的に名前付けされた撮影データから個体差情報込みの神経細胞 3 次元配置図を作成する
2. 名前付けが全くされていない撮影データに対して細胞核自動アノテーション手法を開発する

神経細胞の 3 次元配置図を作成するのは, 自動アノテーションに有用な情報として, 細胞ごとの位置の個体差情報が得られるからである. また, 本研究では人手によるアノテーション支援を行うソフトウェアの改良も合わせて行う. 専門家の手によるアノテーション情報の蓄積は自動アノテーションの精度向上の基礎となるためである.

[研究の着眼点]

個体差情報を持つ精密な神経細胞配置図の作成により細胞名特定の精度向上が期待できる

もし線虫の体位に依存しない細胞の位置に加え、細胞位置の自然な個体差情報を蓄積できれば自動アノテーションの劇的な精度向上が期待できる。なぜなら、特徴量のばらつきは分散分析法が示すように、統計科学における予測や推定に重要な役割を果たすためである。

3. 研究の方法

(1) 個体差情報を有する細胞配置図の作成: 補助的な教師データからの統計的形状モデルの構築

個体差情報を有する細胞配置図は統計的形状モデル (Stegmann et al., 2002) を基礎として構築した。統計的形状モデルは物体の平均形状とその物体の形状のばらつきを考慮した統計モデルで、肺や胃など撮影された臓器の画像領域推定にしばしば応用される。ここでは、物体の形状とそのばらつきを線虫の神経細胞配置とその個体差によるばらつきと考えることで、個体差情報を有する神経細胞配置図の作成に応用可能である。通常、統計的形状モデルの構築には予め個体間で神経細胞の対応関係がとれていることが必要であるが、神経細胞間の対応をとること自体が解くべき課題であり、神経細胞の撮影データから得られた細胞の位置には細胞の対応関係がとれていない。この問題は、専門家による部分的なアノテーションが期待できる場合、確率的成分分析 (PPCA; Bishop et al., 1999) を利用することで解決できる。なぜなら、EM アルゴリズム (Dempster et al., 1977) を利用することで、対応関係のない神経細胞の座標を補完しつつ PPCA の学習が行える、すなわち、統計的形状モデルを構築できるからである。

(2) 自動アノテーション: 神経細胞配置図と実際の配置の個体差を考慮した対応づけ

統計的形状モデル構築後、例えば細胞配置図中の神経細胞群と撮影データ中の細胞群の距離の総和を最小にするような対応関係を求めることで自動アノテーションが可能になる。ここで、統計的形状モデルの構築により得られた細胞配置の個体差情報を最大限利用するために、細胞間の距離として神経細胞配置の個体差によるばらつきを考慮した距離を利用することで高精度での自動アノテーションが期待できる。本研究では実施者が開発した Dependent Landmark Drift (Hirose, 2018) を利用した。

4. 研究成果

研究期間が長期に渡ったため以下前半期・後半期に分けて説明する。

[前半期: 自動アノテーション手法の確立]

4. 1 対象データ

まず研究のきっかけとなったデータから説明する。本研究で使用したデータは 188 個体分の線虫の神経細胞の三次元座標データである。データの特徴としては(1)神経細胞が密集している、(2)個体差、姿勢差がある、(3)部分的に専門化が目視により細胞名を特定している、(4)局所的な位置関係が保存されている、ことがあげられる。図 3 左は(a)が個体番号 44、(b)が個体番号 131 の神経細胞座標データである。緑色の点は全て神経細胞を表しており、濃い緑色の点は専門家によって目視の細胞名特定が行われた細胞を表している。神経細胞が密集していること、個体差、姿勢差があること、そして個体によって細胞名が特定されている部分が異なることがわかる。

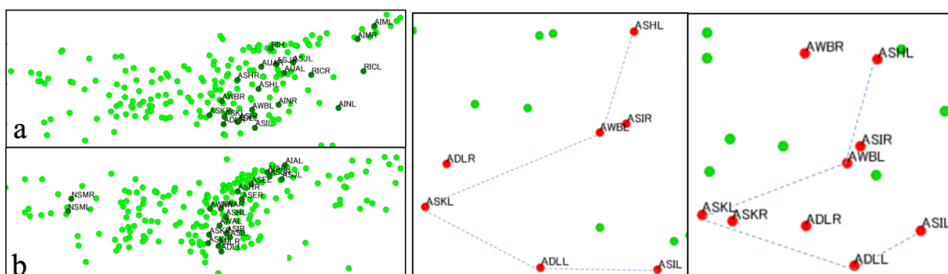


図 3 対象データ (左) とその特徴 (右)。神経細胞の配置の個体差が分かりやすいように一部の細胞を線で結んでいる。

また、図3右は個体番号1と個体番号2でASHL, AWBL, ASKL, ADLL, ASILの位置関係を青い線で結んで表したものである。異なる個体であるが、位置関係が似ていることがわかる。他の神経細胞においても概ね位置関係が保たれるという特徴がある。

4. 4 実験と考察

前述した手法の他に2部グラフマッチングおよび、代表的な点集合位置合わせ手法であるCoherent Point Drift (CPD)を用いて細胞名特定精度の比較を行った。使用したデータは2章で説明したものである。細胞名が特定されている細胞数に対して、細胞名特定に正解した細胞数の比を精度とした。統計的形状モデルを構築する際には観測回数が5回以下の細胞を除いている。そのため統計的形状モデルの神経細胞の種類は127種類となっている。確率的成分分析で統計的形状モデルを構築する際の変動の種類は $K=6$ とした。またDependent Landmark Drift (DLD)のパラメータは $\omega = 0.7$, $\gamma = 10^{-6}$ としている。3つの手法の精度は表1にまとめられている。

表1

	2部グラフ	CPD	統計的 形状モデル
精度	42.91%	42.64%	42.10%

実験の結果、3つの手法で精度に大きな差がないことがわかった。そこでそれぞれの手法における神経細胞ごとの正解率を比較した。図5左がその図である。赤い細胞ほど正解率が高く、正解率が低くなるにつれて青くなるようにプロットしている。(a)が2部グラフマッチング、(b)がCPDを用いた手法、そして(c)が統計的形状モデルを用いた手法の図である。この図から全ての手法で細胞ごとの正解率の傾向が似ていることがわかる。以上より神経細胞ごとに細胞名特定が行いやすさに差があると考えられる。

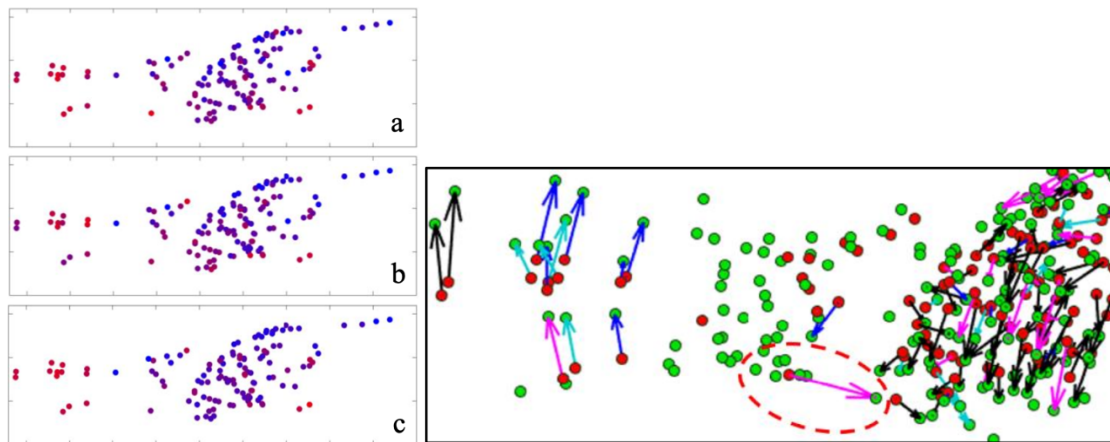


図5

統計的形状モデルの有効性についても考察を行った。図の緑色の点は細胞名未特定データ、赤色の点はアトラスである。矢印は全て細胞同士の対応関係を表しており、青色、桃色、水色、黒色の矢印はそれぞれ、CPDを用いた手法と統計的形状モデルを用いた手法の両方で正解している箇所、統計的形状モデルでのみ正解している箇所、CPDでのみ正解している箇所、両方の手法で正解できなかった箇所を表している。赤い点線で囲まれた部分はRMEDという神経細胞の対応関係を表しているが、RMEDは周囲の細胞と位置関係を保っていないことがわかる。この統計的形状モデルを用いた手法ではRMEDの細胞名特定に成功している。これは統計的形状モデルを用いた手法が周囲の細胞と位置関係を保たない細胞に対しても有効な手法となる可能性を示唆する結果である。

[後半期: 位置合わせ手法の改善]

前半期に開発した細胞名特定手法の核は点集合位置合わせ手法であった。後半期は前記細胞名特定手法の精度をさらに向上させるために、点集合位置合わせ手法の改善に注力した。その結果、機械学習・人工知能分野のトップジャーナルである IEEE Transactions on Pattern Analysis and Machine Intelligence に2報の論文を報告するという大きな成果を得ることができた。以下では、簡単にこれらの内容について報告する。

[1] A Bayesian Formulation of Coherent Point Drift, IEEE TPAMI (Impact Factor =24.314), 43(7): 2269-2286, 2021, Osamu Hirose (単著).

この論文は形状の位置合わせ手法の最も有名な手法の1つである Coherent Point Drift (CPD) [Myronenko et al. IEEE TPAMI 2010] をベイズ統計学の観点から定式化することで、CPD が潜在的に持つ欠点の克服を目指したものである。この定式化により、CPD の基礎的な仮定である モーションコヒーレンスの意義を数理的に明らかにした。また、この定式化を踏まえて CPD を拡張した Bayesian coherent point drift (BCPD) とその高速化手法を提案した。開発手法は CPD の汎用性を犠牲にすることなく (1) 計算時間の短さ, (2) 外れ点に対する耐性, (3) 標的形状の回転に対する耐性の点で CPD を大きく上回った。この論文の意義はベイズ統計学を基礎とする定式化に基づいて CPD をスタート地点から作り直し、理論と実用の両面で発展させた点にある。

[2] Acceleration of Non-Rigid Point Set Registration with Downsampling and Gaussian Process Regression, IEEE TPAMI (Impact Factor =24.314), 43(8): 2858-2865, 2021, Osamu Hirose (単著).

数百万点以上で構成される形状の位置合わせには、以下の[研究業績2]で開発した高速化手法を利用しても相当の計算時間が必要であった。この論文で、提案者の開発手法である Bayesian Coherent Point Drift (BCPD) の高速化手法である BCPD++ を提案した。この手法では、形状の位置合わせを (1) 点を減らすことによる対象形状の情報量縮約, (2) 情報量縮約後の形状に対する位置合わせ, (3) 情報量縮約時に除去された点の位置合わせの3つのステップに分割することで高速化する。その結果、開発手法が1,400万点以上の点を含む形状に適用できることを数値実験により実証した(図3)。この研究の意義は、位置合わせ精度に優れるという長所を持つものの大規模な3D形状への適用が難しいという短所を持つソフトマッチング(SM)法から短所のみを取り除き、大規模な3D形状にSM法が実用できることを実証したという点にある。

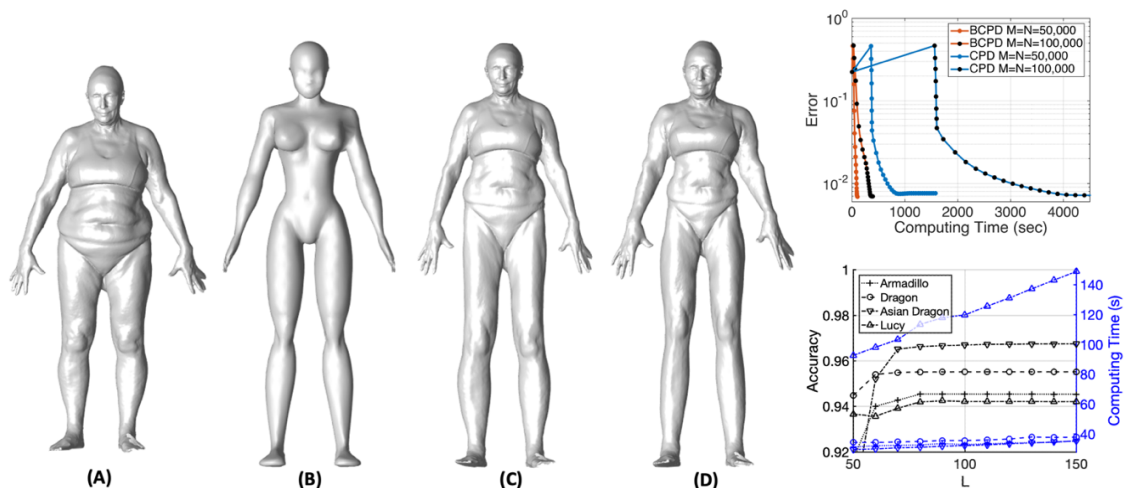


図8. (左)実施者の開発手法である BCPD の応用例. (A) が基準形状, (B) が目標形状, (C), (D) が開発手法を使って基準形状を変形し目標形状に位置合わせしたものである。適用結果の違いは設定パラメータの違いによる。開発手法は基準形状の形状特徴を残したまま目標形状に重ね合わせることができる。(右上) BCPD は標準手法である CPD と比較してはるかに高速であり、同等以上のマッチング精度を達成する。(右下) BCPD++ は BCPD に比べ計算時間を大幅に短縮する。横軸は高速化パラメータを表す。実験に使用された形状は少なくとも 10 万点で構成される。1400 万点以上を含む Lucy データに必要とされる計算時間は入出力を含め約 2 分である。

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 3件/うち国際共著 0件/うちオープンアクセス 2件）

1. 著者名 Osamu Hirose	4. 巻 Early Access
2. 論文標題 Acceleration of non-rigid point set registration with downsampling and Gaussian process regression	5. 発行年 2020年
3. 雑誌名 IEEE Transactions on Pattern Analysis and Machine Intelligence	6. 最初と最後の頁 1 -- 8
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/TPAMI.2020.3043769	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 Hirose Osamu	4. 巻 Early access
2. 論文標題 A Bayesian Formulation of Coherent Point Drift	5. 発行年 2020年
3. 雑誌名 IEEE Transactions on Pattern Analysis and Machine Intelligence	6. 最初と最後の頁 1-18
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/TPAMI.2020.2971687	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 Hirose Osamu, Kawaguchi Shotaro, Tokunaga Terumasa, Toyoshima Yu, Teramoto Takayuki, Kuge Sayuri, Ishihara Takeshi, Iino Yuichi, Yoshida Ryo	4. 巻 15
2. 論文標題 SPF-CellTracker: Tracking Multiple Cells with Strongly-Correlated Moves Using a Spatial Particle Filter	5. 発行年 2018年
3. 雑誌名 IEEE/ACM Transactions on Computational Biology and Bioinformatics	6. 最初と最後の頁 1822 ~ 1831
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/TCBB.2017.2782255	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 1件/うち国際学会 0件）

1. 発表者名 広瀬 修
2. 発表標題 形状位置合わせの高速化と形状合成への応用
3. 学会等名 2021年度統計関連学会連合大会
4. 発表年 2021年

1. 発表者名 広瀬 修
2. 発表標題 近傍点共変動法の一般化および高速化
3. 学会等名 医用画像研究会, 信号処理研究会, バイオメトリクス研究会, 画像工学研究会, 情報センシング研究会, メディア工学研究会 合同研究会 (招待講演)
4. 発表年 2020年

1. 発表者名 広瀬 修
2. 発表標題 点群マッチング問題の正規マルコフ確率場による定式化
3. 学会等名 2019年度統計関連学会連合大会
4. 発表年 2019年

1. 発表者名 広瀬 修
2. 発表標題 統計的形状モデルを利用した大規模点群レジストレーションとその高速化
3. 学会等名 金沢統計シンポ, 融合する統計科学
4. 発表年 2018年

1. 発表者名 広瀬 修
2. 発表標題 形状事前分布を利用した頑健な点群レジストレーション手法の開発
3. 学会等名 2018年度統計関連学会連合大会
4. 発表年 2018年

1. 発表者名 広瀬 修
2. 発表標題 形状事前分布を利用した高速な点群マッチング手法の開発
3. 学会等名 2018年度人工知能学会全国大会
4. 発表年 2018年

1. 発表者名 広瀬 修
2. 発表標題 変位場の滑らかさを前提としない頑健な点群マッチング手法の開発
3. 学会等名 第31回人工知能学会全国大会
4. 発表年 2017年

1. 発表者名 広瀬 修
2. 発表標題 形状事前分布を利用した頑健な点群マッチング手法の開発
3. 学会等名 第105回人工知能基本問題研究会
4. 発表年 2017年

1. 発表者名 広瀬 修
2. 発表標題 形状事前分布を利用した 3D 形状復元アルゴリズムの開発
3. 学会等名 情報処理学会第80回全国大会
4. 発表年 2017年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

Bayesian coherent point drift (BCPD/BCPD++)
<https://github.com/ohirose/bcpd>
金沢大学 Papers of the Month 2021年2月
<https://o-fsi.w3.kanazawa-u.ac.jp/showcase/papers/archive.php>
形状の位置合わせを飛躍的に高速化させるアルゴリズムを発見
<https://www.kanazawa-u.ac.jp/rd/90075>
BCPD/BCPD++: Shape Transfer Demo
<https://youtu.be/9G1oqyNra4Q>
BCPD/BCPD++: Non-rigid Registration Demo
<https://youtu.be/cET6gKAvjw0>
Dependent Landmark Drift
<https://github.com/ohirose/dld>
SPF-CellTracker
<https://github.com/ohirose/spf>

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
--	---------------------------	-----------------------	----

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
---------	---------