

平成 30 年 6 月 6 日現在

機関番号：32612

研究種目：新学術領域研究(研究領域提案型)

研究期間：2013～2017

課題番号：25120014

研究課題名(和文) スパースモデリングを促進する可視化基盤の強化

研究課題名(英文) Consolidation of Visualization Platform Toward Facilitating Sparse Modeling

研究代表者

藤代 一成(Fujishiro, Issei)

慶應義塾大学・理工学部(矢上)・教授

研究者番号：00181347

交付決定額(研究期間全体)：(直接経費) 41,200,000円

研究成果の概要(和文)：本研究は、スパースモデリングにより抽出された高次元データの十数から数十次元の説明変数を、可視化を用いてさらに2,3,4次元まで圧縮し、所与の問題を記述する物理空間におけるデータの挙動を視覚的に理解させる情報空間に変換する基盤技術を構築した。ここでは、そのデータ解析処理に対し明示的に解析者の視覚的フィードバックを取り込み、スパースモデリングに基づく高次元データの解析処理において、人の対話処理が可能となるHuman-in-the-Loopを実現し、データ駆動科学に対し大幅な促進を図ることを目指した。

研究成果の概要(英文)：The goal of this project is to build a visualization platform for understanding the behavior of a given high-dimensional data in physical space by projecting it onto a designated information space. For this purpose, we transform relatively low-dimensional data obtained through sparse modeling to 2, 3, or 4-dimensional space where we can fully visualize the characteristics of the original high-dimensional data. The idea behind the project is to establish an interactive model called human-in-the-loop by incorporating visual feedback from the users in the analysis of high-dimensional data based on sparse modeling.

研究分野：情報学

キーワード：可視化 スパースモデリング 機械学習 トポロジー

### 1. 研究開始当初の背景

さまざまな分野で観測が日常的に実施されている現在、そこから生み出される大量の高次元データに隠されている有益な情報を発見することは重要である。スパースモデリングは、高次元データに内在するスパース性を利用し、計算量が次元数に対して指数的に爆発するような状況でも、実際的な時間でデータから最大限の情報を効率よく抽出できる技術である。しかし、スパースモデリングによって抽出されたデータであっても、通常その説明変数の個数は十数から数十次元にまでしか落ちない。そこで、さらに視覚的に理解可能な2,3,4次元にまで説明変数を圧縮し、所与の問題を記述する物理空間におけるデータの挙動を情報空間に変換して分析可能にするような新たな情報可視化技術が必要になる。

### 2. 研究の目的

本研究では、明示的に解析者の視覚的フィードバックを取り込むことで、スパースモデリングに基づく高次元データの解析において分析者の対話的処理が重要な役割を果たす **Human-in-the-Loop** を実現し、データ駆動科学に対し大幅な促進を図ることを目指す。

図1は、提案する処理基盤を模式的に示したものである【雑誌論文⑦, 学会発表②③】。

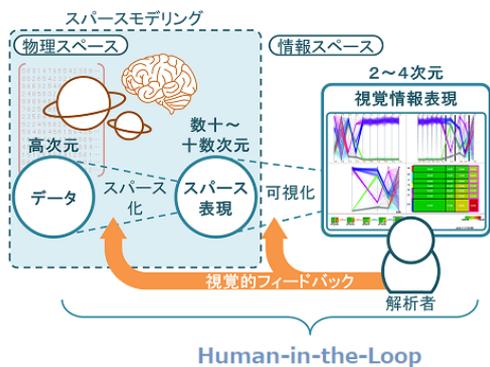


図1 スパースモデリングのための可視化基盤

### 3. 研究の方法

図2に本研究の体制と課題を示す。研究期間の前半は前項の目的を達成するために、以下の三課題を設定した：

- 【課題1】最適商写像選択過程の機械学習
- 【課題2】商写像表現変換のための潜在変数モデリング
- 【課題3】次元圧縮写像の視覚メタファとグリフ設計

そして研究後半は、上記三課題の対合として導出される複合課題A~Cを設定した：

【複合課題A (課題1と2の対合)】商写像推定の高可用性：商写像変換の推定に用いられる学習モデルの複雑度自体をデータから推定する枠組みを構築し、商写像推定の可用性を高めるような拡張を行う。

【複合課題B (課題2と3の対合)】視覚メタファ設計：視覚メタファを適切に設計することで、次元圧縮等に対応するデータの商写像表現の視認性を向上させる。特に高次元から中次元データの視認性向上につながる商写像変換を、データ解析の用途に応じてさまざまな可視化技法と連動させて構築する。

【複合課題C (課題3と1の対合)】履歴学習：可視化の出自管理に注目し、多次元可視化技法群の利用履歴を機械学習し、新たなデータに対して、最適な技法とその利用シナリオを推奨する可視化設計支援機能を実現する。

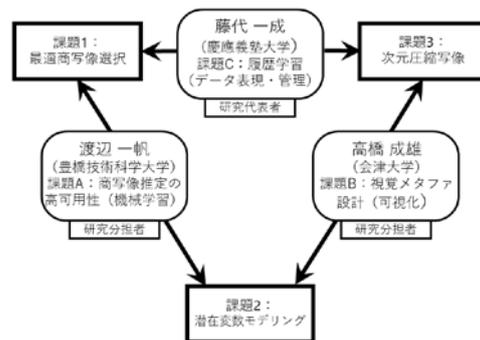


図2 研究の体制と課題

### 4. 研究成果

(1) グラフスペクトル軸縮約平行座標系  
課題1~3の総合的な議論に着手するために多次元データの視覚解析に用いられる代表的な2次元情報可視化技法の一つである、平行座標系表示(図書①)に注目した。この技法は、各変数の値を互いに平行に並んだ軸上にプロットし、各データサンプルを、軸群を横断する折れ線1本で表現する。平行座標系表示は、データの次元数とサンプル数が増加するにつれ、次元軸間が狭くなり折れ線同士が稠密に重なるため、視覚的にデータを把握することが困難になる、視覚的乱雑さの問題が存在する。そこで、グラフスペクトル解析(課題2)によりデータ相関性を捉え、複数の座標軸を順次単一の軸に縮合し、より少ない座標軸を用いて多次元データに内在する大局的な挙動の視覚的理解を促進する手法を開発した(雑誌論文⑭, 学会発表⑱)。

図3に、18次元722サンプルのUSDA食品成分データに本手法を適用した結果を示す。同図上に示すような初期状態から出発した後、数回に渡る段階的な軸併合を経て、同図下に示すような、既知の分析結果に匹敵する次元クラスタリングの結果が得られた。

### (2) 非対称バイクラスタリング

成果(1)に示した軸縮約平行座標系表示は、中次元の多変量データから適切な変数を選択はできるが、軸の縮合後もデータサンプル数は減少せず、多変量データに埋もれた興味ある部分空間を、軸とサンプルの双方から同時に同定する能力に関しては課題を残している。

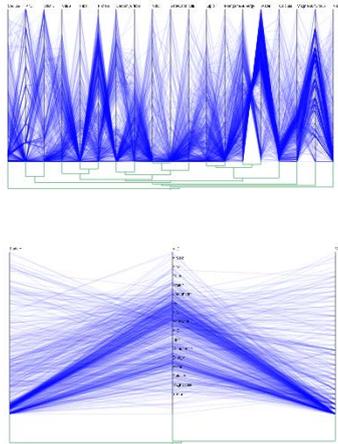


図3 軸縮約平行座標系の視覚効果

この欠点を緩和するために、本研究では非対称バイクラスタリング(課題2)を用いて、有意なデータ相関情報を提供する部分空間を効果的に構成する手法を開発した(学会発表⑤⑩)。ここで「非対称」と冠したのは、クラスタリングに用いる距離尺度が軸とサンプルで相異なる点に由来している。

図4に、成果(1)と同一のUSDA食品成分データに本手法を適用した結果を示す。システムインターフェースは、大きく6つのコンポーネントから構成されている。左上は部分空間探索中の平行座標系表示、右上は太い仕切り線により軸クラスタを明示した平行座標系表示、左下は縮合された平行座標系である。右下は各ブロックのデータ相関性の高さを表す色付きブロック図(低:緑—赤:高)であり、複合課題Bの新たな可視化メタファとして導入したものである。さらに下部には、部分空間探索の過程を組織的に管理する履歴ツリー(複合課題C)とバイクラスタリングの目的関数値も確認される。非対称バイクラスタリングによって、自動的に高い相関性をもつデータサンプルと次元を同時クラスタリングする際に、分析者は色付きブロック図を参照しながら、相関性に乏しい次元や異常値と考えられるデータサンプルを意図的に選択・削除し、より高い相関性をもつ部分空間を効率的かつ効果的に生成することができる。

なお、複合課題Aに対応するために、制約付き von Mises-Fisher モデルに基づくベイス推論を用いて、例えば軸・データの各クラスタ数の初期値を適切に決定できるように手法を増強した(雑誌論文⑫)。

さらに、この非対称バイクラスタリング環境を、本研究が所属する新学術領域研究「スパースモデリング」の天文学班が取り扱ってきた、Ia型超新星とよばれる天体の分類問題にも適用し、天文学者が長期間の議論の末、経験的に得ていた従来の分類と概ね一致する結果を半自動的に得ることに成功しており、同領域研究を代表する成果の一つとして認定されている(雑誌論文⑬)。

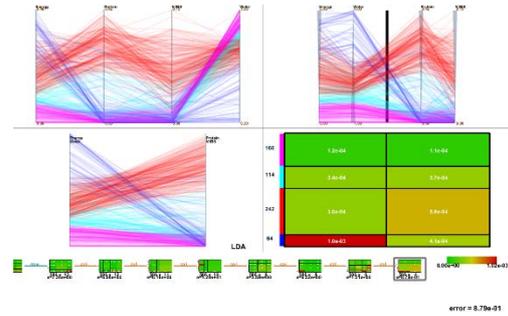


図4 非対称バイクラスタリング環境

### (3) TimeTubes

時間依存の多次元データに対するスパースモデリングを支援するために、本研究ではTimeTubesとよばれる、成果(1)(2)とは異なる可視化メタファ(複合課題B)を開発し(雑誌論文①),「スパースモデリング」の天文学班に属する広島大学宇宙科学センターで長期間観測されてきたブレーザーの7次元偏光データ分析に適用した(雑誌論文③⑩)。図5に示すように、TimeTubesは、各時刻の観測データを、偏光パラメータを中心、その誤差を径長とする楕円に変換し、データ信頼度を同心状の不透明度分布で表現した後、隣接断面間を自由曲線で補間することにより、ボリュームメトリックなチューブ形状を構成し、観測光の強度や色はチューブの視覚属性としてマッピングする(学会発表⑩)。

観測データの誤差と観測の欠損によって生じる観測データ特有の不確実性に対処するために、同一ブレーザーに関する複数の観測所からのデータを視覚的に融合することで一覧性を増強し、より詳細で正確なデータ検証を可能にした(複合課題A)。実際、広島大学の観測データに、アリゾナ大学がウェブサイトで公開している観測データを視覚的に融合することに成功している(学会発表④)。さらに、光度が急激に上昇する現象であるフレアや、偏光角が大きく回転する現象であるローテーションのような、特徴的な現象を含むブレーザー時系列データの部分時空間を学習し、類似検索する機能(複合課題C)も実現した(学会発表①)。

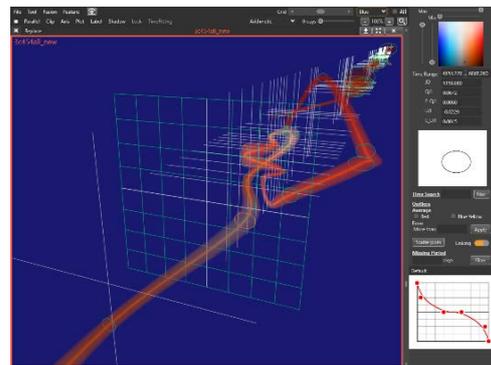


図5 TimeTubes

#### (4) オンライン学習の精度評価

オンライン学習は、逐次的にデータが与えられる状況において、適応的に予測を変化させていく枠組みである。オンライン学習における逐次予測は、データセットの出方についての最悪値を最小化するミニマックス問題として定式化され、その精度に基づいた学習モデル選択法の有効性が多くの問題において報告されている。課題1や複合課題Cにおける商写像選択の履歴の学習はオンライン学習として定式化することができ、モデル選択は課題2や複合課題Aにおいてクラスタ数や次元数などの潜在変数モデルを選択する際に必要となる。しかしながら、一般にミニマックス解は計算困難であるため、多くの場合において効率的な近似が必要とされている。そこで、学習データ数が大きい極限においてミニマックス解を精度良く近似する予測法を開発した(雑誌論文⑮, 学会発表⑱)。

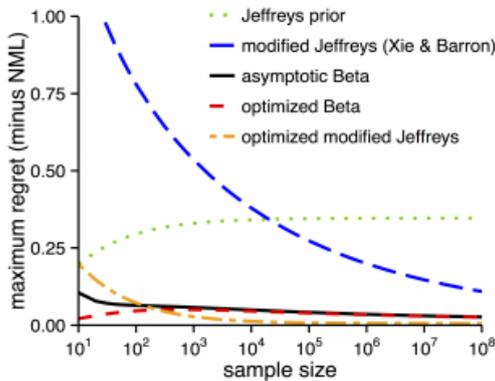


図6 データ数に対するオンライン学習の精度

#### (5) クラスタ数推定とレート歪み関数

課題1での商写像選択過程の学習や課題2での商写像変換の設計では、学習モデルの複雑度自体をデータから推定する機構が必要とされる(複合課題A)。クラスタリング等に幅広く用いられ、混合数をデータから自動的に決定することができる、無限混合モデルの学習を考察した。過学習が起りやすい最尤推定による学習法と、別の既存法であるカーネルベクトル量子化法を統一的に扱った学習法を提案し、適切な設定のもとで過学習を抑制できることを示した(雑誌論文⑲)。また、混合モデルの学習による情報圧縮の限界はレート歪み関数により与えられることが情報理論において知られている。無限混合モデルから得られるクラスタリング法とレート歪み関数の関係を明かにした(雑誌論文④)。レート歪み関数の評価が与えられているのは、ごく少数の例に限られている。歪み尺度を柔軟に設計することのできるカーネル法に基づく歪み尺度を含め、種々の歪み尺度に対し、レート歪み関数の評価を与えた(雑誌論文⑤⑥⑧)。

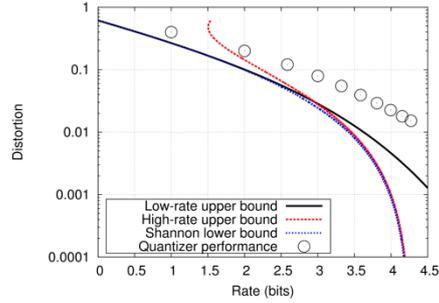


図7 レート歪み関数の評価

#### (6) 多変量データの微分トポロジー解析

課題2の商写像変換の設計の一手法として本研究では、多変量データを関数のサンプルデータとみなし、定義域のサンプル点に関して複数の関数値の挙動を可視化する手法を開発した。その基本アイデアは、関数値の逆像(ファイバー)の変化を追跡することであり、従来可視化分野で標準的手法として用いられていた単変量関数の微分トポロジー解析手法を、多変量関数を対象に高度化したことにある(雑誌論文⑩)。

本手法においては、特に3次元定義域から2次元値域への関数サンプルデータに着目し、ファイバーのトポロジー変化の型を分類することで、多変量関数データを視覚的記号表現へと変換している。その点において本研究の成果は、複合課題BやCへの橋渡しの意味をもつ。また、単変量関数データの解析(例えば学会発表⑥)に関する知識に基づき、成果(1)や(2)とは異なる知見を提供できる道具としても利用可能である。

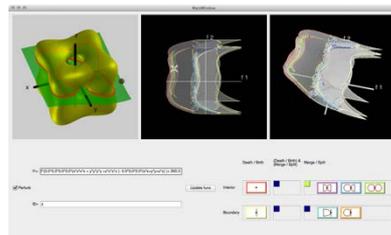


図8 多変量データの微分トポロジー解析

#### (7) ネットワークを介した視覚データ解析

多変量データの可視化を行う場合、データサンプル間に陰に陽に存在する幾何学のおよび論理的な隣接関係は、データに潜在する特徴の理解だけでなく、データをスクリーン空間上に適切に配置する際の重要な手がかりを与える。本研究では、課題3および複合課題Bに対応するため、ネットワークを介したデータ解析及び視覚メタファ設計を継続的に実施した。

具体的には、ネットワークを介した高次元データの視覚的な次元削減(学会発表⑰⑳)、注釈付きインフォグラフィックス表現(学会発表⑪)およびその対称性に基づく強調処理



図9 高次元画像空間におけるカテゴリ分類化

(学会発表⑧), ノード注釈ラベルの重なりを回避するネットワーク可視化手法 (雑誌論文②, 学会発表⑨⑩), ノード間の半順序ネットワーク可視化のための最適化 (雑誌論文⑨) などの成果を得た. また, 関連する講演活動 (学会発表⑦⑫⑬) などを行った.

##### 5. 主な発表論文等

[雑誌論文] (計 23 件, すべて査読有)

- ① I. Fujishiro, N. Sawada, M. Nakayama, H.-Y. Wu, K. Watanabe, S. Takahashi, M. Uemura: TimeTubes: Visual exploration of observed blazar datasets, To appear in *Journal of Physics: Conference Series*, 2018.
- ② H.-Y. Wu, S. Takahashi, R. Ishida: Overlap-free labeling of clustered network based on Voronoi tessellation, *Journal of Visual Language & Computing*, Vol. 44, pp. 106-119, 2018 [DOI: 10.1016/j.jvlc.2017.09.008].
- ③ M. Uemura, R. Itoh, I. Liodakis, D. Blinov, M. Nakayama, N. Sawada, H.-Y. Wu, I. Fujishiro: Optical polarization variations in the blazar PKS1749+096, *Publications of the Astronomical Society of Japan*, Vol. 69, No. 6, pp. 96:1-96:12, 2017 [DOI: 10.1093/pasj/psx111].
- ④ 小林 真佐大, 渡辺 一帆: デリクレ過程平均法のレートひずみ理論による解釈, 電子情報通信学会論文誌 A, Vol. J100-A, No. 12, pp. 475-486, 2017
- ⑤ K. Watanabe: Rate-distortion bounds for kernel-based distortion measures, *Entropy*, Vol. 19, No. 7, id. 336, 2017 [DOI: 10.3390/e19070336].
- ⑥ K. Watanabe: Projection to mixture families and rate-distortion bounds with power distortion measures, *Entropy*, Vol. 19, No. 6, id. 262, 2017 [DOI: 10.3390/e19060262].
- ⑦ 藤代 一成, 高橋 成雄, 渡辺 一帆, H.-Y. Wu: スパースモデリングと情報可視化, 電子情報通信学会誌, Vol. 99, No. 5, pp. 466-470, 2016
- ⑧ K. Watanabe, Shiro Ikeda: Rate-distortion functions for gamma-type sources under absolute-log distortion measure, *IEEE Transactions on Information Theory*, Vol. 62, No. 10, pp. 5496-5502, 2016 [DOI: 10.1109/TIT.2016.2602100].
- ⑨ H.-Y. Wu, S. Takahashi, H. N. Miyamura, S. Ohzahata, A. Nakao: Inferring partial orders of nodes for hierarchical network layout, *Journal of Imaging Science and Technology*, Vol. 60, No. 6, 2016 [DOI: 10.2352/J.ImagingSci.Technol.2016.60.6.060407].
- ⑩ M. Uemura, R. Itoh, L. Xu, M. Nakayama, H.-Y. Wu, K. Watanabe, S. Takahashi, I. Fujishiro: TimeTubes: Visualization of polarization variations in blazars, *Galaxies*, Vol. 4, No. 3, Article No. 23, 2016 [DOI: 10.3390/galaxies4030023].
- ⑪ D. Sakurai, O. Saeki, H. Carr, H.-Y. Wu, T. Yamamoto, D. Duke, S. Takahashi: Interactive visualization for singular fibers of functions  $f: \mathbb{R}^3 \rightarrow \mathbb{R}^2$ , *IEEE Transactions on Visualization and Computer Graphics*, Vol. 22, No. 1, pp. 945-954, 2016 [DOI: 10.1109/TVCG.2015.2467433].
- ⑫ K. Watanabe, H.-Y. Wu, S. Takahashi, I. Fujishiro: Asymmetric biclustering with constrained von Mises-Fisher models, *Journal of Physics: Conference Series*, 699, 012028, 2016 [DOI: 10.1088/1742-6596/699/1/012018].
- ⑬ M. Uemura, K. S. Kawabata, S. Ikeda, K. Maeda, H.-Y. Wu, K. Watanabe, S. Takahashi, I. Fujishiro: Data-driven approach to Type Ia supernovae: Variable selection on the peak luminosity and clustering in visual analytics, *Journal of Physics: Conference Series*, 699, 012009, 2016 [DOI: 10.1088/1742-6596/699/1/012009].
- ⑭ 能野 琴, H.-Y. Wu, 渡辺 一帆, 高橋 成雄, 藤代 一成: グラフスペクトル解析を用いた平行座標系の軸縮約, 画像電子学会誌, Vol. 44, No. 3, pp. 447-456, 2015
- ⑮ K. Watanabe, T. Roos: Achievability of asymptotic minimax regret by horizon-dependent and horizon-independent strategies, *Journal of Machine Learning Research*, Vol. 16, pp. 2357-2375, 2015.
- ⑯ K. Watanabe, S. Ikeda: Entropic risk minimization for nonparametric estimation of mixing distributions, *Machine Learning*, Vol. 99, No. 1, pp. 119-136, 2015 [DOI: 10.1007/s10994-014-5467-7].

[学会発表] (計 72 件)

- ① 澤田 奈生子, 中山 雅紀, 植村 誠, 藤代 一成: TimeTubes: 特徴検索によるブレーザー観測データ解析の詳細化と効率化, 第 284 回画像電子学会研究会, 広島大学東千田キャンパス (広島県広島市), 2018 年 3 月 1-2 日
- ② I. Fujishiro: Visual analysis tools for facilitating sparse modeling, Joint Keynote for *ACM SIGGRAPH Asia 2017 Symposium on Visualization and Symposium on Education*, Bangkok, Thailand, November 27-30, 2017. (基調講演)

- ③ I. Fujishiro: Facilitating sparse modeling with sophisticated visual analysis tools, *The First Chinese-Japanese Joint Visualization Workshop*, Beijing, China, July 24, 2017. (招待講演)
- ④ N. Sawada, M. Nakayama, H.-Y. Wu, M. Uemura, I. Fujishiro: TimeTubes: Visual fusion and validation for ameliorating uncertainties of blazar datasets from different observatories, *CG International 2017*, Yokohama, Japan, June 27-30, 2017.
- ⑤ H.-Y. Wu, Y. Niibe, K. Watanabe, S. Takahashi, M. Uemura, I. Fujishiro: Making many-to-many parallel coordinate plots scalable by asymmetric biclustering, *IEEE PacificVis 2017*, Seoul, Korea, April 18-21, 2017.
- ⑥ Y. Takeshima, S. Takahashi, I. Fujishiro: Parameter-setting for topology-accentuated transfer function for volume raycasting, *TopoInVis 2017*, Tokyo, Japan, February 27-28, 2017.
- ⑦ S. Takahashi: Schematizing graphics for visual understanding, *The 8th International Conference on Graphics and Image Processing*, Tokyo, Japan, October 29-31, 2016. (招待講演)
- ⑧ K. Yasuda, S. Takahashi, H.-Y. Wu: Enhancing infographics based on symmetry saliency, *VINCI 2016*, Dallas, USA, September 24-26, 2016.
- ⑨ R. Ishida, S. Takahashi, H.-Y. Wu: Adaptive blending of multiple network layouts for overlap-free labeling, *iV2016*, Lisbon, Portugal, July 19-22, 2016.
- ⑩ L. Xu, M. Nakayama, H.-Y. Wu, K. Watanabe, S. Takahashi, Makoto Uemura, I. Fujishiro: TimeTubes: Design of a visualization tool for time-dependent, multivariate blazar datasets, *SAS NICOGRAPH International 2016*, Hangzhou, China, July 6-8, 2016.
- ⑪ Hsiang-Yun Wu: Focus+context metro map layout and annotation, *Spring Conference on Computer Graphics 2016*, Slovenia, Slovakia, April 27-29, 2016 (Best paper award, 2nd best presentation award).
- ⑫ S. Takahashi: Egocentric design in geospatial mapping, *IEEE BDVA2015*, Hobart, Tasmania, Australia, September 22-25, 2015. (招待講演)
- ⑬ S. Takahashi: Formulating optimization models for visual computing, *Smart Graphics 2015*, Chengdu, China, August 26-28, 2015. (招待講演)
- ⑭ Y. Ohtaka, S. Takahashi, H.-Y. Wu, N. Ohta: Using mutual information for exploring optimal light source placements, *Smart Graphics 2015*, Chengdu, China, August 26-28, 2015.
- ⑮ R. Ishida, S. Takahashi, H.-Y. Wu: Interactively uncluttering node overlaps for network visualization. *iV2015*, Barcelona, Spain, July 21-24, 2015.
- ⑯ K. Watanabe, H.-Y. Wu, Y. Niibe, S. Takahashi, I. Fujishiro: Biclustering multivariate data for correlated subspace mining, *IEEE PacificVis 2015*, Hangzhou, China, April 14-17, 2015.
- ⑰ Y. Gao, H.-Y. Wu, K. Misue, K. Mizuno, S. Takahashi: Visualizing bag-of-features image categorization using anchored maps, *VINCI 2014*, Sydney, Australia, August 5-8, 2014.
- ⑱ K. Nohnno, H.-Y. Wu, K. Watanabe, S. Takahashi, I. Fujishiro: Spectral-based contractible parallel coordinates, *iV2014*, Paris, France, July 15-18, 2014.
- ⑲ A. Barron, T. Roos, K. Watanabe: Bayesian properties of normalized maximum likelihood and its fast computation, *IEEE ISIT 2014*, Honolulu, USA, June 29-July 4, 2014.
- ⑳ M. Kazuyo, H.-Y. Wu, S. Takahashi: Manipulating bilevel feature space for category-aware image exploration, *IEEE PacificVis 2014*, Yokohama, Japan, March 4-7, 2014.
- [図書] (計 1 件)
- ① 藤代 一成, 齋藤 隆文, 乃万 司, 高橋 成雄, 馬場 雅志, 宮田 一乗 (編著): 改訂新版 コンピュータグラフィックス, 画像情報教育振興協会, 2015 (総 441 頁)
- [その他]
- ① 文部科学省科学研究費補助金新学術領域研究 (平成 25 年度~29 年度)「スパースモデリングの深化と高次元データ駆動科学の創成」ホームページ <http://sparse-modeling.jp/>
- ② 慶應義塾大学理工学部藤代研究室ホームページ <http://fj.ics.keio.ac.jp/>
6. 研究組織
- (1) 研究代表者  
藤代 一成 (FUJISHIRO Issei)  
慶應義塾大学・理工学部・教授  
研究者番号: 00181347
- (2) 研究分担者  
渡辺 一帆 (WATANABE Kazuho)  
豊橋技術科学大学・工学研究科・講師  
研究者番号: 10506744
- 高橋 成雄 (TAKAHASHI Shigeo)  
会津大学・コンピュータ理工学部・教授  
研究者番号: 40292619
- (3) 研究協力者  
吳 湘筠 (WU Hsiang-Yun)  
ウィーン工科大学・ビジュアルコンピューティングと人間中心技術研究所・研究員