

[共同研究報告]

機械のためのビジュアライゼーションによるインタラクション分析

堂園翔矢

久保田晃弘

1. はじめに

何らかの方法で、鑑賞者自身が作品体験の一部として参加し、さらにその参加の仕方によって、作品自体の形態や状態が変化するような作品を、インタラクティブな作品と呼ぶ。インタラクティブな作品の起源は、サイバネティックス⁽¹⁾の考え方が、美術作品の制作に取り入れられたことにある⁽²⁾。サイバネティクスという語は、舵とりが上手であることを意味する、ギリシア語の「kybernetikos」に由来する。生物や機械における制御や恒常性を実現する、サイバネティックスの考え方を美術作品に取り入れるためには、作品が鑑賞者や環境の情報を「入力」し、それに対して作品内部で何らかの「処理」を行い、その結果を、作品の外観や動きを通じて鑑賞者に「出力」する機能を持つ必要がある。そうすることで、作品＝鑑賞者の系（システム）がフィードバックループを形成し、作品と鑑賞者のインタラクティブな関係が生まれる。

本研究は、このサイバネティックな作品の核である「インタラクション」の本質を、理論的、実験的に解明することを目的としている。そのための具体的な事例として、1998年に本学に設立された多摩美術大学情報デザイン学科で教鞭をとっていた、故三上晴子教授（1961～2015）⁽³⁾が残したインタラクティブ・メディア・インスタレーション作品のひとつである《Eye-Tracking Informatics》（以下《ETI》）⁽⁴⁾を取り上げる。《ETI》は、山口情報芸術センター（YCAM）の委嘱作品として、2011年に制作された。三上の没後の2018年にYCAMによって再制作され、翌2019年に多摩美術大学アートテークギャラリーで再展示された⁽⁵⁾。同年2019年には、NTTインターコミュニケーション・センター（ICC）でも長期展示された⁽⁶⁾。さらに2023年には、√K Contemporary ギャラリーで特別展示が行われた⁽⁷⁾。

《ETI》は「視ることを視る」をテーマにした、インタラクティブなメディア・インスタレーション作品である。鑑賞者の視線をデバイスによって検出し（入力）、その視線の動きからアルゴリズムに3次元の仮想構造体が生成され（処理）、それが鑑賞者に再び提示される（出力）。この動的に構造化された視線を、鑑賞者が再び視ることで「視ることを視る」という、セカンド・オーダー・サイバネティックスのシステムが生まれる。さらに、同時に体験している、あるいは以前参加した別の鑑賞者の視線を重ねて提示することで、鑑賞者同士の視線によるコミュニケーションが可能になる（図①）。

著者らのグループは、2018年に設立された多摩美術大学アートアーカイブセンター（以下 AAC）の収蔵資料のひとつである「三上晴子アーカイブ」を活用して、これまで三上の作品^(8~11)や教育活動に関する、さまざまな制作研究を行ってきた。なか

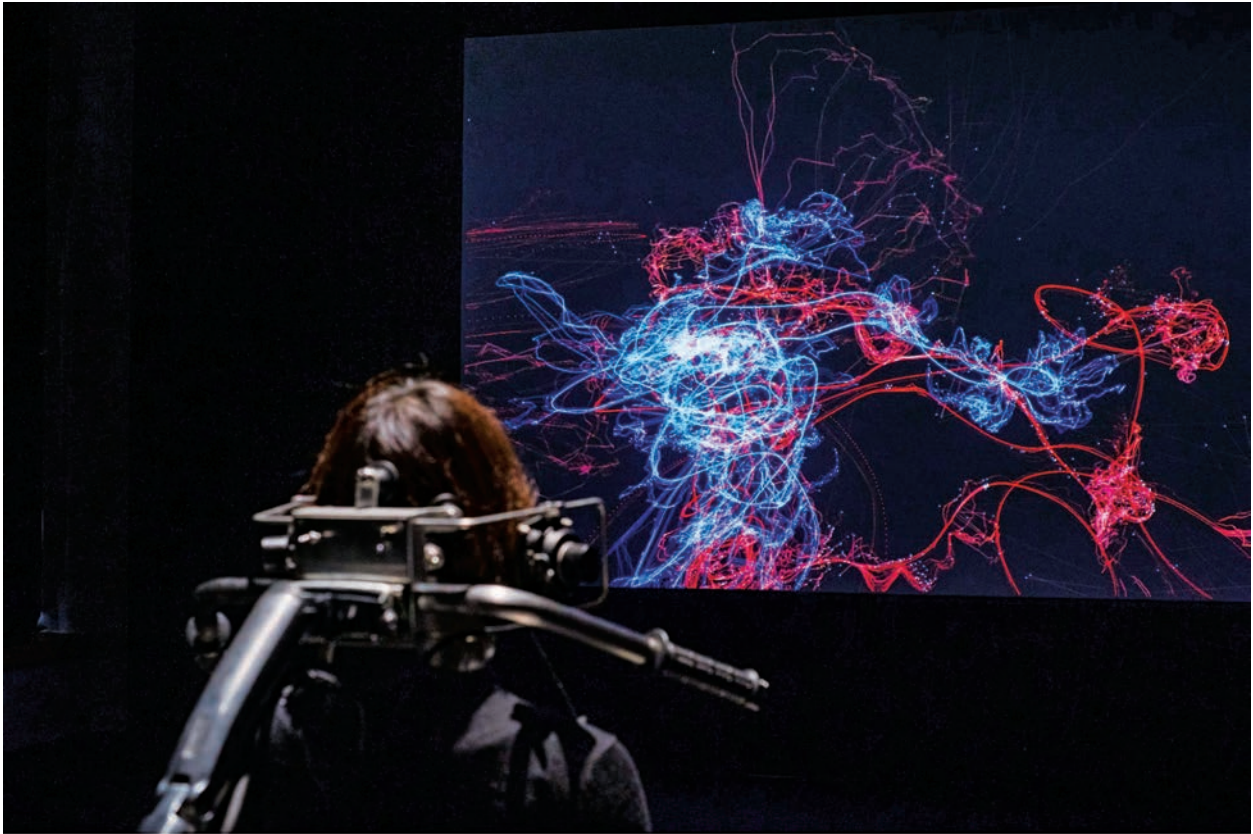


図1 三上晴子《Eye-Tracking Informatics Version 1.1》—— YCAM との共同研究成果展，撮影：竹久直樹

でも、《ETI》が NTT [ICC] で展示された際の 2020 年の 1 月 12 日から 2 月 28 日までの期間、作品のプログラムの平川紀道と ICC の協力を得て取得した、770 の視線および生成された 3 次元仮想構造体のデータを用いて、いくつかのインタラクション分析を行ってきた。具体的には、まず個別のデータに対して、

- ・取得した視線データとその速度や加速度の可視化
- ・生成された 3 次元仮想構造体の可視化
- ・視線の動きと仮想構造体の形状の相関の分析

を行い、そこからいくつかのパターンやモードを抽出した。代表的なものとして、視線があまり動かず、滑らかな構造が生まれる「グライディングモード」と、視線が小まめに動き、くると縮れた構造をとる「ターンモード」の 2 つがある。さらにこのモードは、仮想空間のなかを視線が散策する「探索モード」と、生成された構造を追いかけたり逃げたりする「追跡(逃避)モード」に分けられる(図 2)⁽¹²⁾。

この個別データの分析を踏まえ、次に取得したデータの全体像を把握するために、《ETI》の全取得データの機械学習による解析と分類を行った。視線の動きと仮想構造体の形態データ

を、Variational Auto Encoder (変分オートエンコーダー、以下 VAE) という、データの特徴を獲得するための教師なしモデルを用いてデータ全体を分類し、インタラクションの典型例や特殊例を抽出した⁽¹³⁾。こうしたミクロとマクロのインタラクション分析を通じて、《ETI》におけるインタラクションの全体像が少しずつ見えてきた。

本稿では、こうした《ETI》におけるインタラクションの初期分析の成果を踏まえて、機械学習によるインタラクション分析の、より解像度の高い方法を提案、検討した結果を報告する。

2. 機械学習によるインタラクション分析

本研究のインタラクション分析に用いている VAE は、エンコーダ=デコーダ構造を有した、ニューラルネットワークによる教師なし学習モデルのひとつである⁽¹⁴⁾。VAE はディープラーニングによる生成モデルのひとつで、学習データの特徴を捉えて分類するだけでなく、学習データセットに似たデータを生成することができる。エンコーダによって、膨大な学習データを低次元の潜在空間 (Latent space) に圧縮することで、特徴量の抽出が可能になるだけでなく、潜在空間上の座標を逆にデコードすることで、新たなデータを生成することができる。



図② 3次元仮想構造体の各モード

VAEの実装方法には、全結合（Dense）、畳み込みニューラルネットワーク（CNN）、再帰型ニューラルネットワーク（RNN）がある。《ETI》の体験時の視線データと3次元構造体の座標データという時系列データを分析するにあたって、これまでは、各回の体験時のデータ全体（5分の体験ごとに約15000点のサンプル）を1次元の静的な配列として解析する全結合を用いていた。しかしながら、この方法はデータに既知の構造がある場合、ノード（人工ニューロン）間の結合が過剰になってしまい、例えば《ETI》のインタラクションにおける、時間的な側面を学習結果に反映することがむずかしくなる。

音楽のような時系列データをニューラルネットワークを用いて学習、分類した先行事例として、徳井直が率いるQosmo社の《AI DJ Project》⁽¹⁵⁾がある。ここでは、音楽の楽曲という時系列的なデータの特徴を抽出、分類するために、人間の知覚特性に合わせた周波数の対数-時間軸グラフであるメル周波数スペクトログラム画像を用い、それを画像認識でよく用いられるCNNによって解析している⁽¹⁶⁾。CNNとは、その名前の通り「畳み込み」という操作を加えたニューラルネットワークである。画像認識における畳み込みとは、元の画像に空間フィルターを掛け合わせて、その積分値によって出力画像を得る操作のことである。この操作によって、元画像の局所的な特徴を際立たせた特徴マップを作成し、それを次元削減によって分類する。《ETI》のインタラクション分析においては、生成された仮想構造体全体の形を分析するだけでなく、個別データの分析の際に得られた、さまざまなパターンやモード、すなわちインタラクションの局所的な特徴量に基づく解析も必要である。そこで以下、この方法をインタラクションの分析に用いることを検討する。

3. 機械のためのビジュアライゼーション

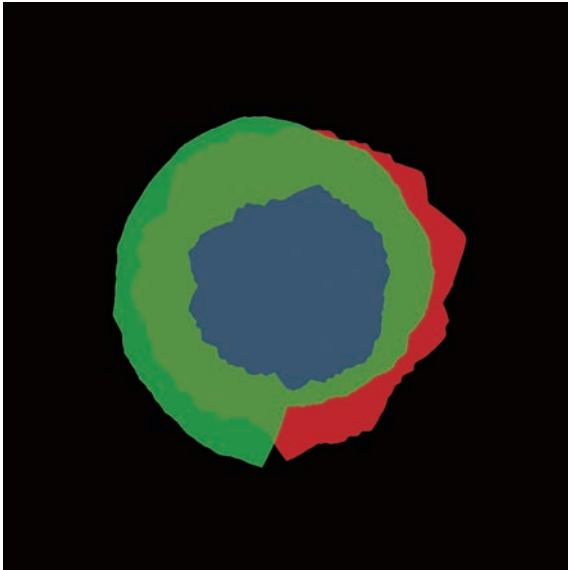
音楽解析の方法を、《ETI》のインタラクション分析で用いる3次元構造体に応用するためには、音楽解析の時のメル周波数スペクトログラムに対応する、何らかの画像データを作成する必要がある。この画像データは、3次元構造体のxyz座標データを可視化したものであるが、それは、人間のための、つまり人間にとってわかりやすい可視化というよりも、むしろ機械学習のアルゴリズムが分類するための、つまり機械にとってわかりやすいビジュアライゼーションである必要がある。

そこで、3次元構造体の時系列座標データxyzそれぞれの座標を、機械が見分けやすくするために、RGBの3色で表現する（x：R、y：G、z：B）。局所的な特徴の出現時間が任意であることと、その時間スケールが重要であるため、一回の体験で一周する極座標表現を用いる。そうすることで、xyzの3つの変数を分離したまま解析することが可能になる。

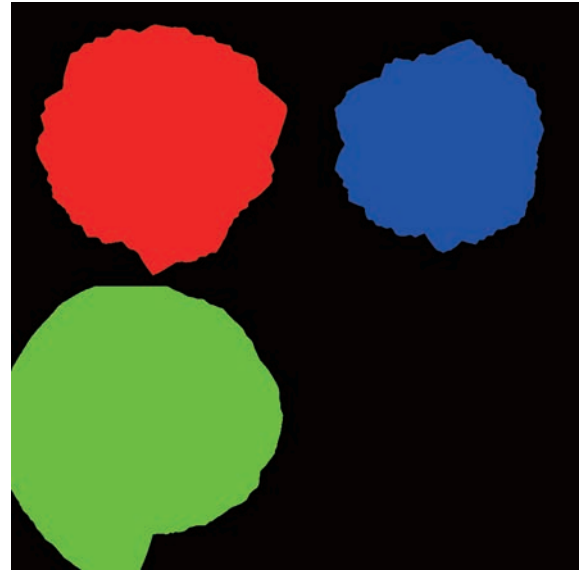
今回は、ビジュアライゼーションの手法として、以下の4つのケースを実験、比較した（図③～⑥）。

- ・CASE 1：RGBを重ねて表示（アルファブレンド）
- ・CASE 2：RGBを重ねて表示（スクリーンブレンド）
- ・CASE 3：RGBの輪郭のみを表示（スクリーンブレンド）
- ・CASE 4：RGBを分けて表示

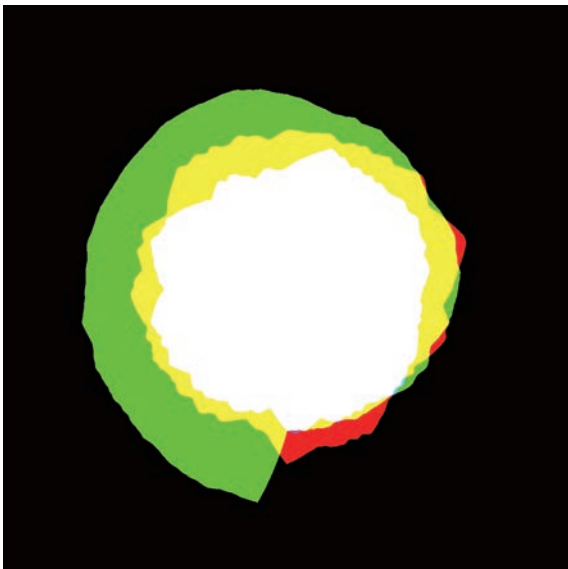
なお、一般に仮想構造体の始点と終点は一致しないため、両者のズレが特徴として認識され難くなるよう、座標データの始点と終点を線形補間によってスムーズに繋いだ。また、今回の解析は仮想構造体に対してのみ行った。仮想構造体の軌跡は基本的に、視線の動きを仮定の加速度として描かれるため、視線データの積分値のようなものと捉えられる。同時に、動きの仮想質量によって、ノイズのスムージング効果も得られる。両者には相関があるため、どちらか一方を解析すれば、インタラクションの特徴が抽出できると考えられる。



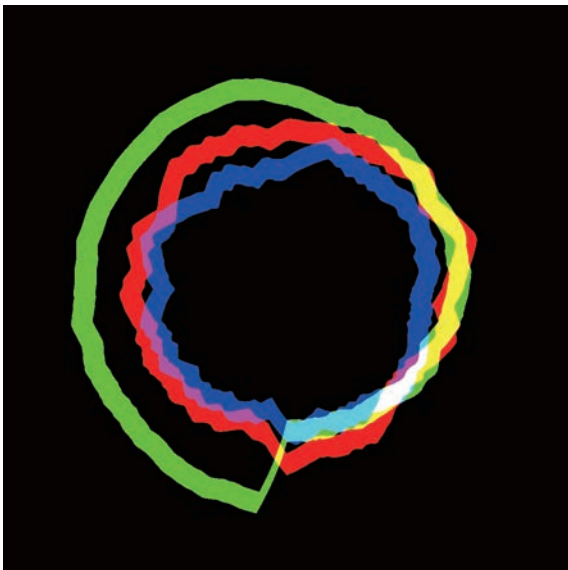
図③ CASE 1 : RGB を重ねて表示 (アルファブレンド)



図④ CASE 4 : RGB を分けて表示



図⑤ CASE 2 : RGB を重ねて表示 (スクリーンブレンド)



図⑥ CASE 3 : RGB の輪郭のみを表示 (スクリーンブレンド)

4. 機械学習による分類結果

本稿では、上記の方法で可視化した画像データから、2014年にOxford大学のVGG (Visual Geometry Group) が発表した、VGG-16というモデルで特徴量を抽出し、それをUMAP (Uniform Manifold Approximation and Projection) を用いて次元削減し分類した。VGG-16は、その名前の通り、16層の深さで構成されるニューラルネットワークで、3×3の小さなフィルターを用いた畳み込み層、プーリング層、全結合層から成っている。さらにそこに、100万枚を超えるイメージを学習させた、事前学習済みのネットワークでもある。また、2018年に発表されたUMAP⁽¹⁸⁾は、トポロジカルデータ解析と呼ばれる数学理論に基づいたもので、データセットを高次元空間上の連続的な構造 (多様体) として捉え、これを低次元空間に射影することで次元削減を行う。UMAPは、これまで使用してきたt-SNE (t-distributed Stochastic Neighbor Embedding) による次元削減方法に比べて、埋め込み次元数によらず実行時間がほぼ一定、かつデータの局所的な構造と大域的な構造の両方を保持することができるため情報の損失が少ない、という特徴を持ち、大規模なデータセットや高次元データの次元削減に、より有効であるとみなされている。

図⑦は、「CASE 3」(図⑥)と「CASE 4」(図④)の画像を、VGG-16モデルによって特徴抽出した例である。一番左の画像が、最初のプーリング層の出力で、その次の画像が、2番目のプーリング層の出力である。プーリング層とは、畳み込み層の後に適用される、入力データを圧縮 (ダウンサンプリング) してより扱いやすい形に変形するための層である⁽¹⁹⁾。この図から、2番目のプーリング層の特徴量を使うとうまく画像を分類でき

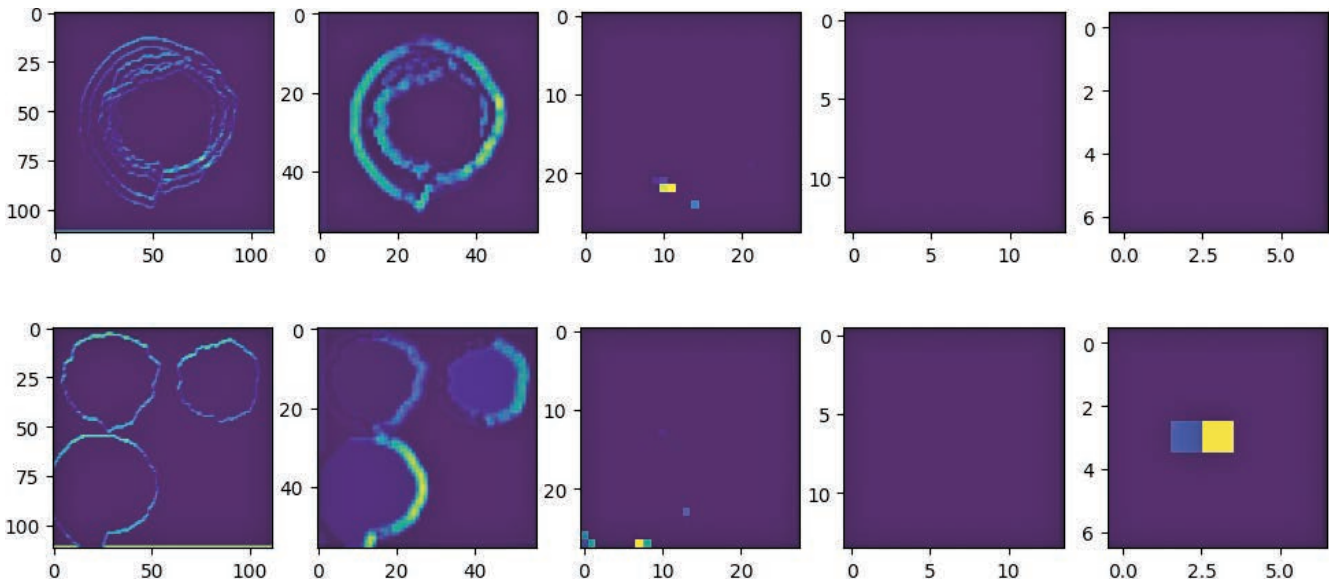


図7 VGG-16 モデルによる画像の特徴マップ (上が CASE 3, 下が CASE 4)

ると判断し、以下の分析ではこの 3136 次元ベクトルの特徴量を用いる。

「CASE 1~4」の 4 種類の方法で仮想構造体の座標を可視化した画像の特徴マップを、UMAP によって次元削減した結果(潜在空間)を、図8~11に示す。

比較のため、以前行った座標データの全結合解析と t-SNE⁽¹³⁾により得られた潜在空間を図12に示す。

全結合と t-SNE を用いた分析結果は、全体がひとつのクラスターにまとまっていて、トラッキングエラーを含む特異な事例を抽出して除外したり、クラスターの中央値としての典型的

な事例を抽出することには有効であった。それに対して、今回の機械のためのビジュアライゼーションと UMAP を用いた方法(図8~11)はいずれも、潜在空間上のプロットが一点に集まることなく特定の領域内に分散し、インタラクションの中にある傾向や分類を発見するのに有効であると考えられる。

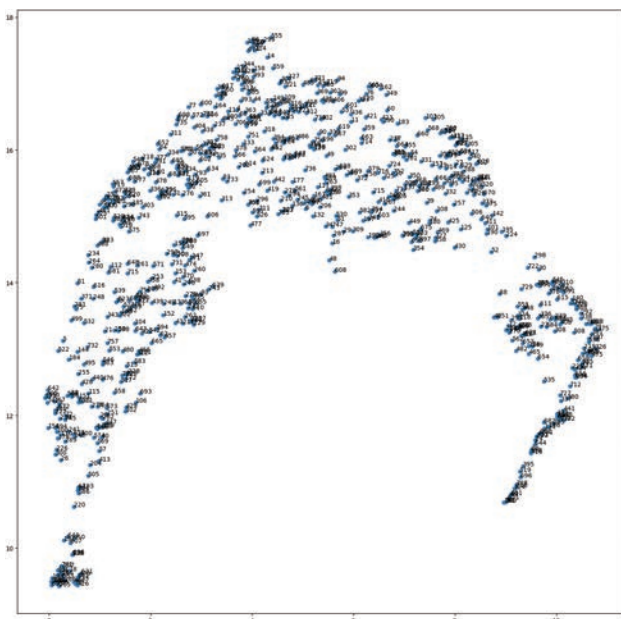


図8 CASE 1 の潜在空間

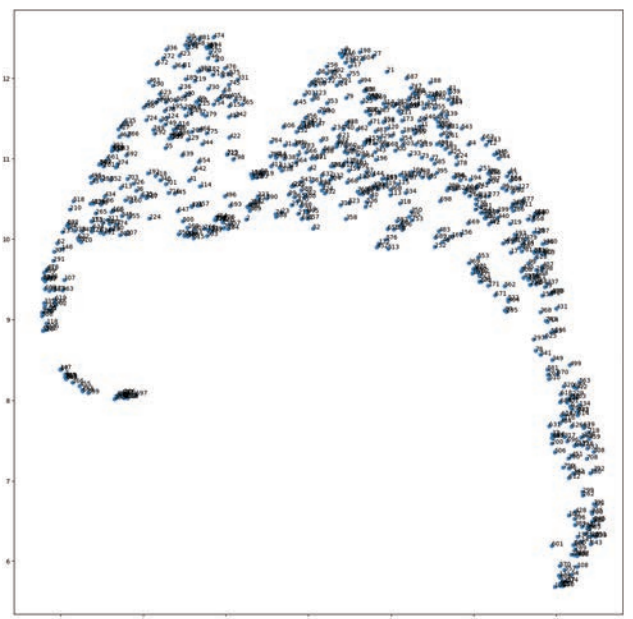


図9 CASE 2 の潜在空間

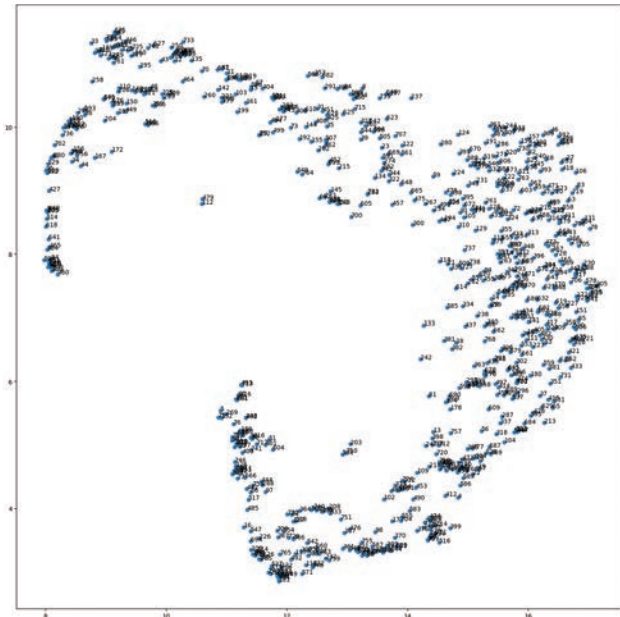


図10 CASE 3の潜在空間

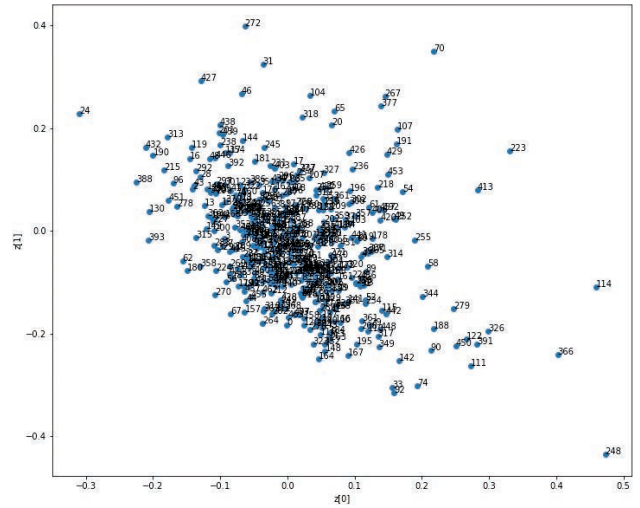


図12 全結合解析による潜在空間

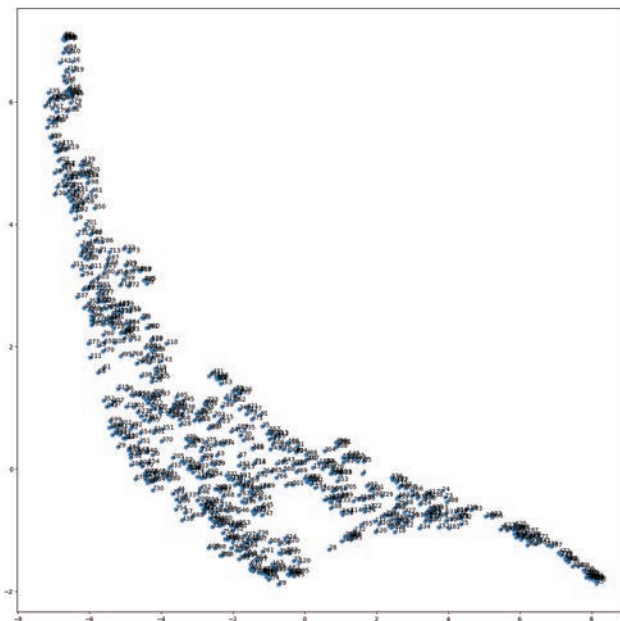


図11 CASE 4の潜在空間

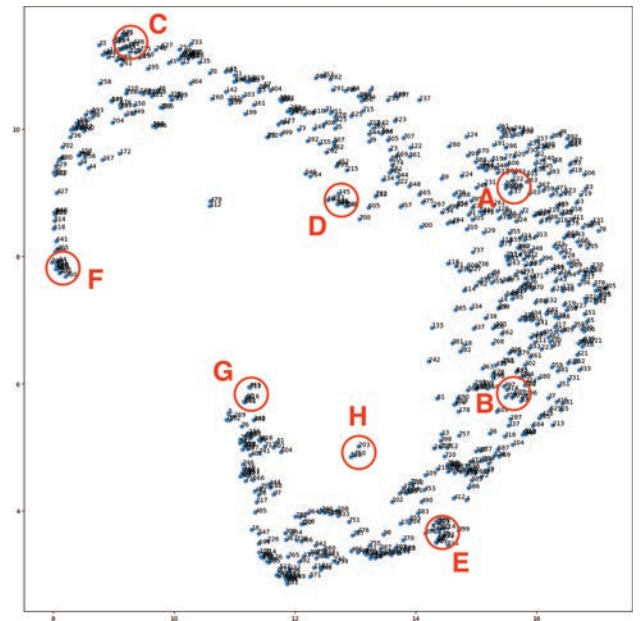


図13 CASE 3の潜在空間の内部構造

5. 潜在空間の内部構造

次に、今回の方法のなかから、RGBの輪郭のみをスクリーンブレンドで重ねて表示した「CASE 3」を取り上げて、その潜在空間の内部構造を詳しくみていく。

まずはじめに、図13のAとBの領域、すなわち潜在空間における分布の内部に位置する可視化画像の例を、図14、15に示す。いずれも、緑(y軸)の時間的変化があまり大きくない、比較的類似した画像が集まっていることがわかる。

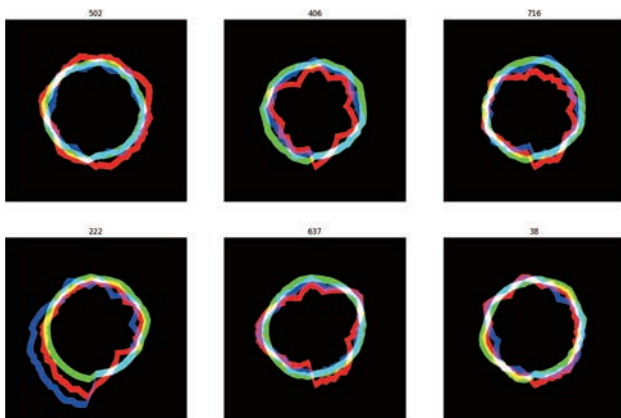


図14 領域Aの可視化画像

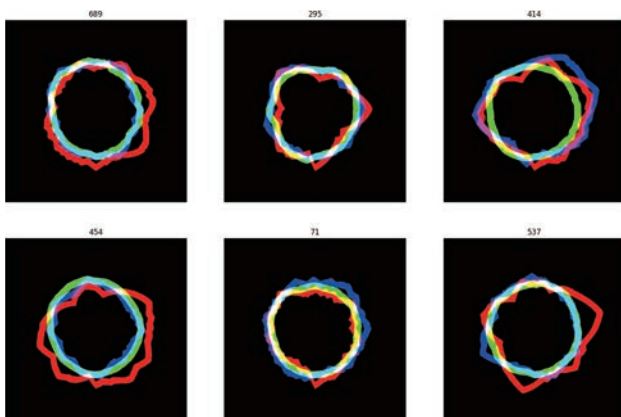


図15 領域Bの可視化画像

また、各領域の仮想構造体の形状のいくつかを、図16、17に示す。赤(x軸)の変動がばらついていることから、厳密な形状の類似性は認められないものの、非常に多様な視線データの中から、全体として類似した構造体が分類されていることがわかる。

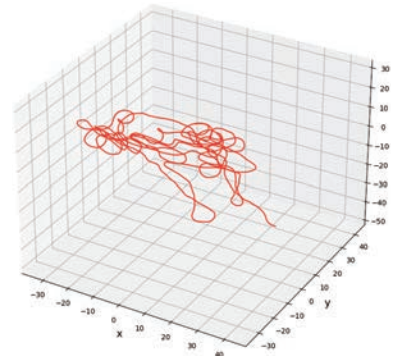
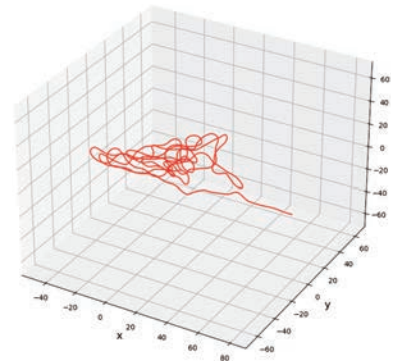


図16 領域Aの仮想構造体

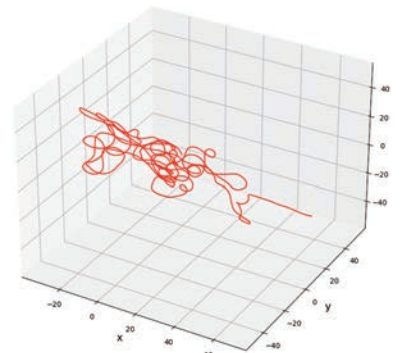
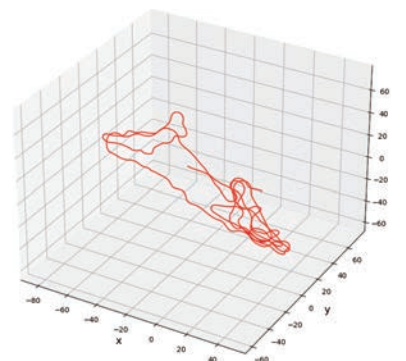


図17 領域Bの仮想構造体

次に分布の境界領域にあたる、図13のC、D、Eの領域をみる。この領域における可視化画像を、図18~20に示す。

領域Eでは、赤(x軸)の形に特徴があるように見える。いずれの例においても、それぞれの領域固有の特徴を、可視化画像からも、仮想構造体の形状からも見て取ることができる。

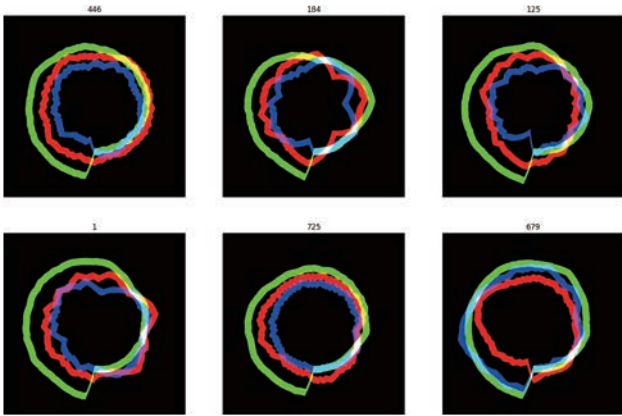


図18 領域Cの可視化画像

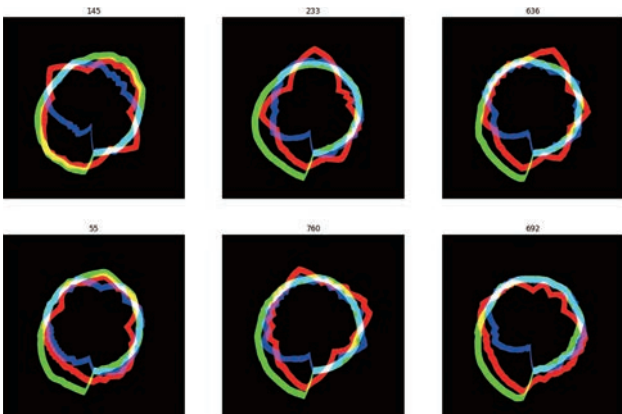


図19 領域Dの可視化画像

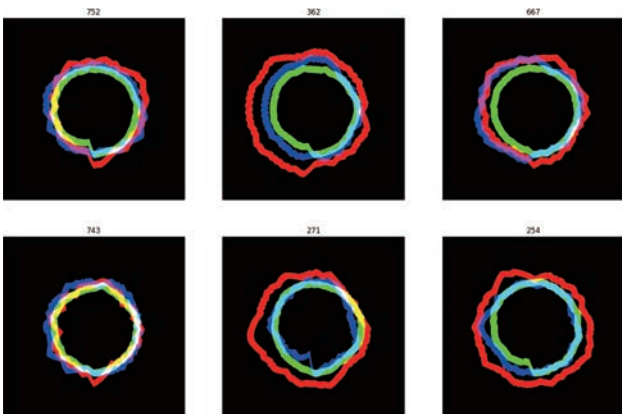


図20 領域Eの可視化画像

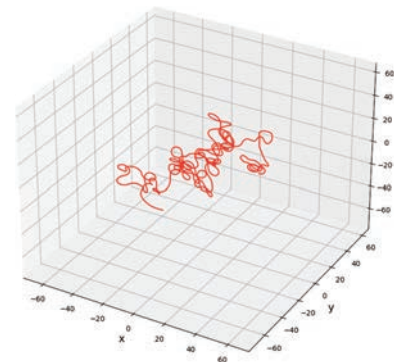
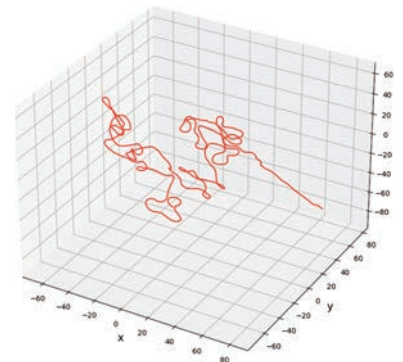


図21 領域Cの仮想構造体

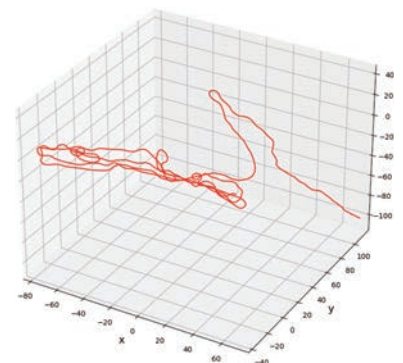
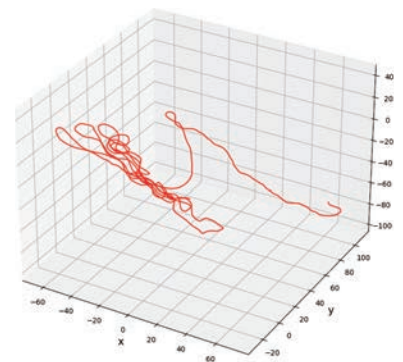


図22 領域Dの仮想構造体

領域CとDに特徴的なのは、緑(y軸)の始点と終点のズレである。この領域の仮想構造体の例を、図21,22に示す。合わせて、領域Eの仮想構造体の例を、図23に示す。

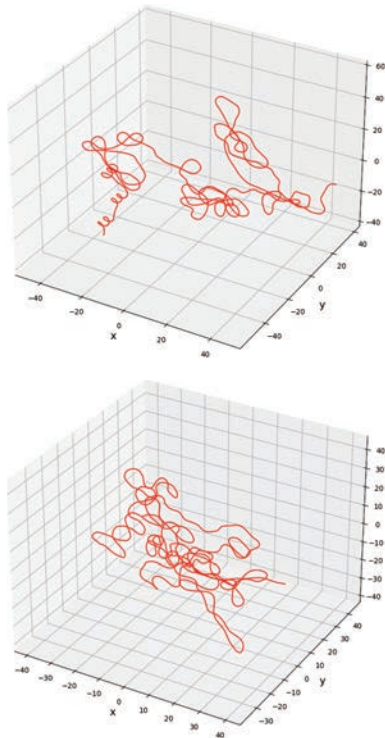


図23 領域Eの仮想構造体

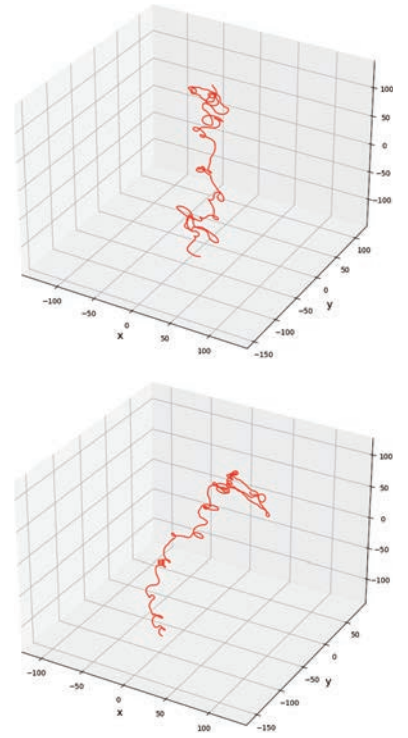


図24 領域Fの仮想構造体

分布の端にあたる、図13のF、Gの領域の可視化画像と、仮想構造体の例を、図24～27に示す。

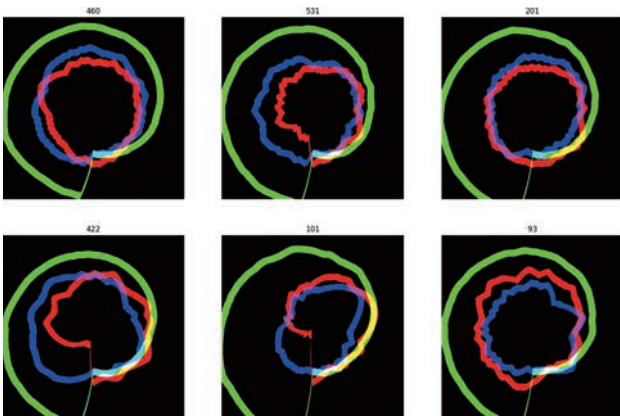


図24 領域Fの可視化画像

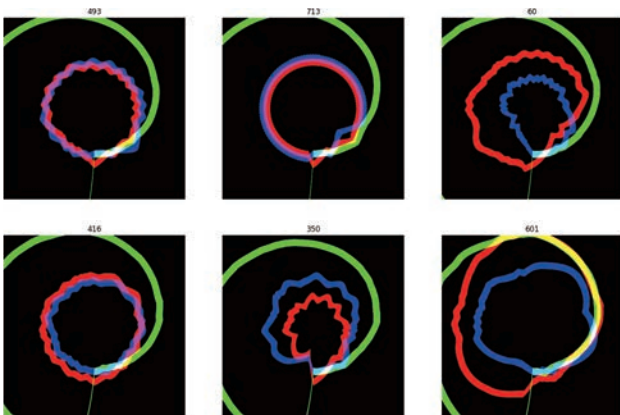


図25 領域Gの可視化画像

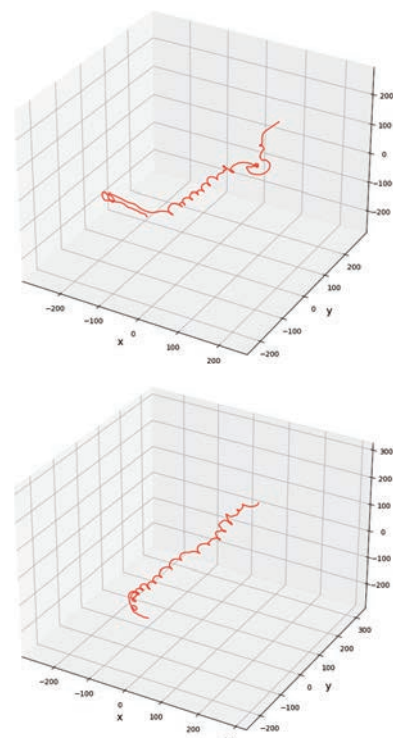


図27 領域Gの仮想構造体

これらはいずれも、以前「特殊例」と分類した、独特の仮想構造体の事例になっている。

最後に孤立した3つの点からなる、領域Hの可視化と仮想構造体をみる(図28,29)。この領域も、領域F、G同様に、緑(y軸)の始点と終点が大きすぎた、特殊な事例となっている。

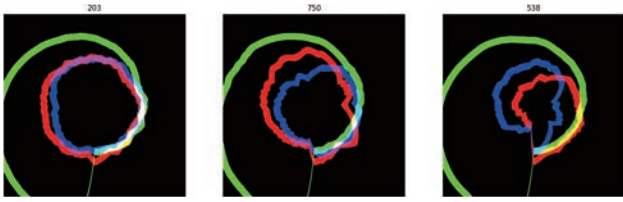


図28 領域Hの可視化画像

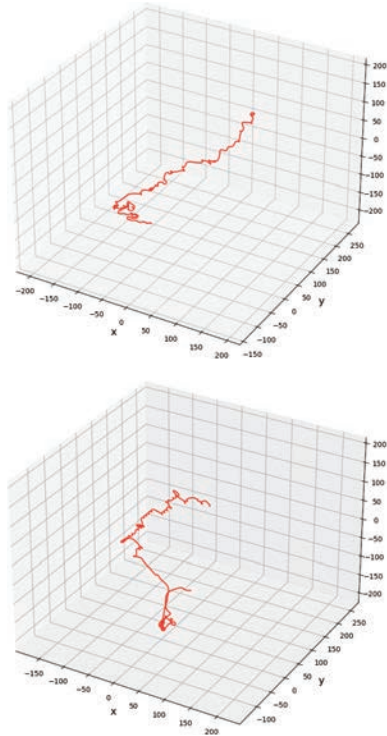


図29 領域Hの仮想構造体

6. ビジュアリゼーションの違いによる比較

座標データのビジュアリゼーションの違いによる、機械学習による分類結果の比較を行うために、RGB (xyz 座標) を分けて表示した「CASE 4」による分類結果 (潜在空間の内部構造) を確認する。

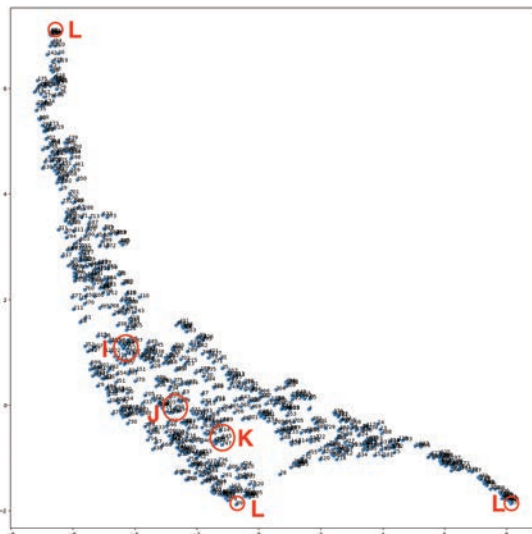


図30 「CASE 4」の潜在空間の内部構造

まず、「CASE 4」による潜在空間 (図30) の分布内部の領域 I、J、K の可視化画像と仮想構造体の例をみる (図31~32)。

いずれも、典型的な形状の構造体であるが、たとえば領域 J では、特徴的な赤 (x 軸) の形状がみられるなど、それぞれの領域ごとに類似した傾向の仮想構造体が集まっている、「CASE 3」の場合同様に、適切に分類が行われていることがわかる。

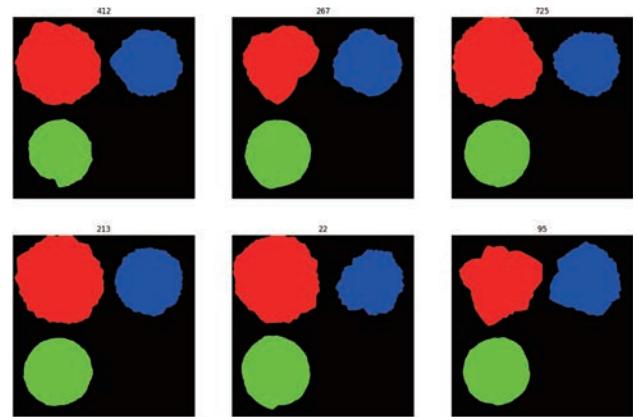


図31 領域Iの可視化画像

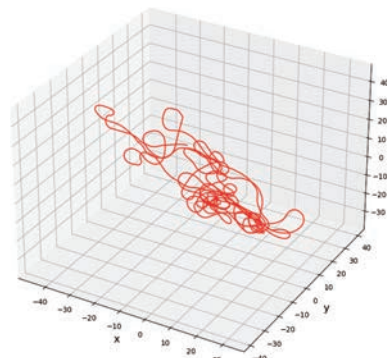
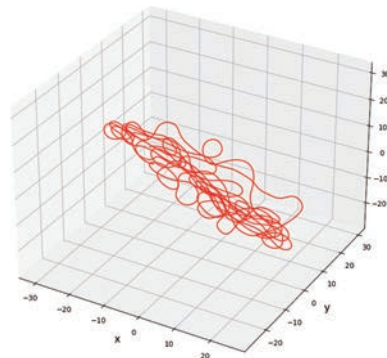


図32 領域Iの仮想構造体

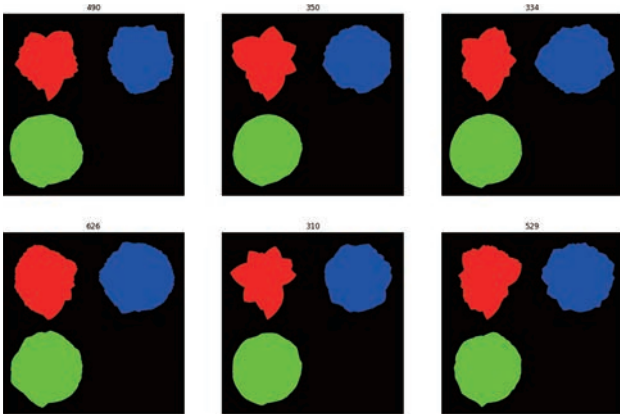


図38 領域Jの可視化画像

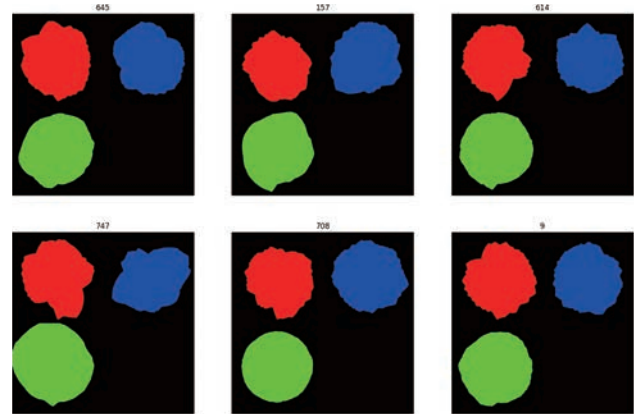


図39 領域Kの可視化画像

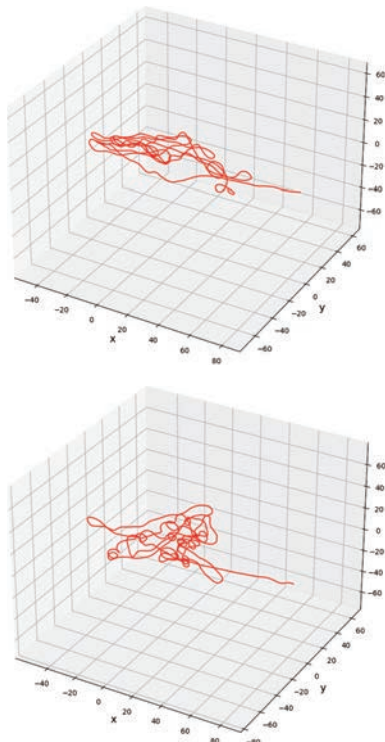


図40 領域Jの仮想構造体

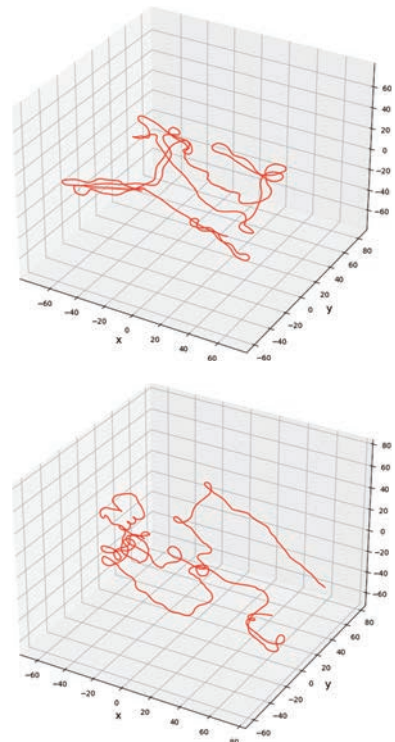


図41 領域Kの仮想構造体

同様に、領域の上端 (98)、右端 (45)、下端 (89) の3点 (領域L) のデータを見ると (図47-38)、これまでの分析と同様に、それらが特殊例に対応していることがわかる。



図47 領域Lの可視化画像

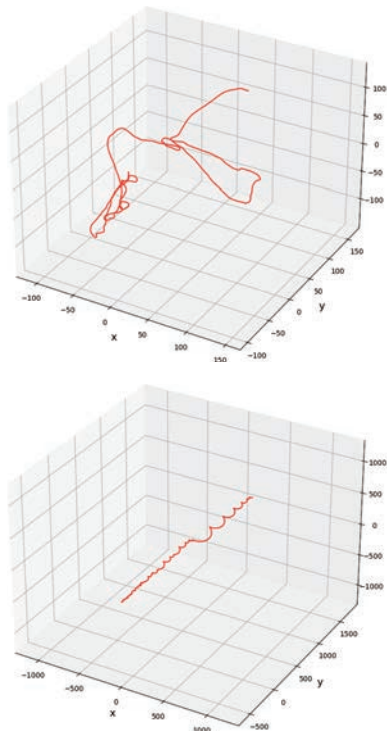


図45 領域Lの仮想構造体（上端 98と右端 45）

7. おわりに

本稿では、VAEを用いたインタラクション分析に、VGG-16+CNNとUMAPという、大規模画像認識のための畳み込みネットワークを活用するため、機械学習に適したデータ・ビジュアライゼーションの提案と、そのいくつかの事例と結果を紹介した。ビジュアライゼーションの具体的な方法としては、インタラクションによって生成された3次元仮想構造体の形状であるxyz座標データを、それぞれRGBの3色に対応させ、各体験事例ごとに一周する極座標表現を採用した。

今回提案した方法を、これまでの全結合を用いた分析と比較したところ、インタラクションの大まかな傾向を反映した、より適切な分類が可能になることがわかった。しかしながら、機械のためのビジュアライゼーションは、人間の美学やわかりやすさとは異なる基準で行う必要があり、可視化における形や色の使用法とその有効性について、今後も継続的に検討を重ねていく必要がある。さらに、この機械のためのビジュアライゼーションは、インタラクション分析に限らず、一般のデータ解析にも応用できると考えられる。そうしたビジュアライゼーションと機械学習のコラボレーションによる、分野横断的な可能性や拡張性についても、検討を行っていきたい。

今回のビジュアライゼーションは、仮想構造体の形状を、絶対（空間固定）座標を用いて表現したことや、構造体をひとつの形状として取り扱ったことにより、前回同様インタラクションの大規模な全体構造についての解析となった。今後、構造体の形状を相対的で動的なローカル座標によって可視化することで、

個別のモードの抽出や、インタラクションのマイクロ構造の分析を行うことができると考えられる。今日用いられている、さまざまな（人間のための）ビジュアライゼーションの手法が、機械学習による分析にどのように活用できるのかを検討することは、ビジュアライゼーションの非人間的活用という広い意味で、多様な制作研究テーマを含んでいると思われる。

同時に、特徴マップの次元削減の方法についても、近年深層学習（自己教師付き学習）とトポロジカルデータ解析（TDA）を組み合わせた、Deep-TDA⁽²⁰⁾という新たな手法が提案されるなど、日夜研究開発が進展している。こうした技術的動向をフォローしながら、インタラクションという極めて複雑で多様な人間と人工物の相互作用についての分析と理解を深め、インタラクティブな作品の本質を、サイバネティックに探る旅を続けていきたい。

謝辞

本研究を進めるにあたっては、共同研究者でアーティストの平川紀道氏、および多摩美術大学美術学部情報デザイン学科メディア芸術コース研究室、アートアーカイブセンターの関係者のみなさんに、多大な協力をいただいた。ここに深く感謝したい。本研究は2022年度学内共同研究「インタラクティブ・アートの理論とアーカイブ」の支援を受けて行われた。関係各位に厚く御礼を申し上げたい。

註

- (1) ウィーナー『サイバネティクス——動物と機械における制御と通信』池原止戈夫他訳、1962年、岩波書店。
- (2) Reichardt, J. (ed.), *Cybernetic Serendipity: The Computer and the Arts*, Studio International Foundation, 1968. <https://archive.org/embed/cybernetic-serendipity> (accessed Sep. 1, 2023).
- (3) 馬定延、渡邊朋也〔編〕『SEIKO MIKAMI 三上晴子 記録と記憶』NTT出版、2019年。
- (4) 三上晴子『Eye-Tracking Informatics』2011年、<https://www.ycam.jp/archive/works/eye-tracking-informatics/>（最終閲覧2023年9月1日）。
- (5) 「三上晴子『Eye-Tracking Informatics Version 1.1』——YCAMとの共同研究成果展」、2019年、<https://aac.tamab.ac.jp/research/reports/2019/1394.html>（最終閲覧2023年9月1日）。
- (6) 三上晴子『Eye-Tracking Informatics』[2011/19]、2019年、<https://www.nttic.or.jp/ja/archive/works/eye-tracking-informatics-2/>（最終閲覧2023年9月1日）。
- (7) 三上晴子「Eye-Tracking Informatics」特別展示、2023年、https://root-k.jp/exhibitions/seiko-mikami_eye-tracking-informatics/（最終閲覧2023年9月1日）。
- (8) 馬定延、久保田晃弘「メディアアートのための生成するアーカイブ試論（前編）」『多摩美術大学研究紀要』第31号、2016年、多摩美術大学、94-98頁。
- (9) 馬定延、久保田晃弘「メディアアートのための生成するアーカイブ試論（中編）」『多摩美術大学研究紀要』第32号、2017年、多摩美術大学、65-76頁。
- (10) 平川紀道、渡邊朋也、馬定延、久保田晃弘「メディアアートのための生成するアーカイブ試論（後編）」『多摩美術大学研究紀要』第33号、2018年、

71-79 頁。

- (11) 久保田晃弘、馬定延、渡邊朋也「メディアアートと生成するアーカイヴ」『軌跡』 vol. 1、2019 年、多摩美術大学アートアーカイヴセンター、10-31 頁。
- (12) 久保田晃弘「メディアアートにおけるインタラクションの分析方法 三上晴子《Eye-Tracking Informatics》を事例として」『軌跡』 vol. 3、2020 年、多摩美術大学アートアーカイヴセンター、25-37 頁。
- (13) 久保田晃弘、平川紀道、堂園翔矢「インタラクションのアーカイヴ」『軌跡』 vol. 4、2021 年、多摩美術大学アートアーカイヴセンター、38-45 頁。
- (14) Kingma, D. P., Welling, M. *Auto-Encoding Variational Bayes*, 2013, <https://arxiv.org/abs/1312.6114> (accessed Sep. 1, 2023).
- (15) Qosmo, *AI DJ Project—The Documentary*, 2022, <https://youtube.com/eEhJuoWCxhw> (accessed Sep. 1, 2023).
- (16) 徳井直「AI DJ 技術解説——9/4「2045」at OKAZAKI LOOPS #2」2016 年、<http://naotokui.net/2016/09/2045-okazakiloops-2/>(最終閲覧 2023 年 9 月 1 日)。
- (17) Simonyan, K. Zisserman, A., *Very Deep Convolutional Networks for Large-Scale Visual Recognition*, https://www.robots.ox.ac.uk/~vgg/research/very_deep/(accessed Sep. 1, 2023).
- (18) UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, <https://umap-learn.readthedocs.io/>(accessed Sep. 1, 2023).
- (19) 等々力康弘「ディープラーニング (Deep Learning : CNN)」OctOpt 技術ブログ、<https://blog.octopt.com/deep-learning-cnn/>(最終閲覧 2023 年 9 月 1 日)。
- (20) Kibardin, E. *Why you should use Topological Data Analysis over t-SNE or UMAP?*, 2023 <https://datarefiner.com/feed/why-tda> (accessed Sep. 1, 2023).