

令和 6 年 6 月 17 日現在

機関番号：13904

研究種目：挑戦的研究（萌芽）

研究期間：2021～2023

課題番号：21K19822

研究課題名（和文）複雑な動作を操る特徴空間の教示学習

研究課題名（英文）Feature space learning for manipulating complicated motions

研究代表者

栗山 繁（Kuriyama, Shigeru）

豊橋技術科学大学・工学（系）研究科（研究院）・教授

研究者番号：20264939

交付決定額（研究期間全体）：（直接経費） 4,800,000円

研究成果の概要（和文）：CGキャラクタを直観的に操作するための特徴空間の教示学習法として、様々な歩容動作と手指操作を対応づける教示方法に取り組み、動きの周波数毎の強度と位相情報を抽出できるネットワークを導入して、手指の動きと全身動作を柔軟かつ安定に関連付ける機構を開発し、一定の効果を確認した。次に、発話音声で教示手段とみなした場合の身振り動作を、特徴量のパターン照合のみにより生成する手法を開発した。粗い解像度では音声に対する同期と生成動作の多様性を獲得し、詳細な解像度では素材データの特徴を復元する手法を開発した。また、データ構造を上下半身に分離し下半身に対しては音声の影響を除外することで、運動学的な誤差を軽減できた。

研究成果の学術的意義や社会的意義

仮想人物の身体動作を対話的に操作する手段として、動作の位相特徴量に着目した要素分解による手指の動きとの対応の教示学習を試みたものの、手指の動きの限界やその再現の困難さ等が原因で、表現に富んだ自然な動きを安定に学習させるには至らなかった。この結果を踏まえて、教示の手段を音声に変更し、身振り動作とは個別に計算する特徴量間の距離を柔軟に統合してパターン照合することにより、両者の時間同期と運動を安定して教示できたことは学術的な意義がある。また、この手段を音声以外にも拡張することにより、様々な方法を用いた深層学習を介さない身体動作の教示法の可能性を示せた点はエンタテインメント分野での波及効果が期待できる。

研究成果の概要（英文）：We developed an example-based learning for intuitively manipulating CG characters using hand movements. We focused on controlling various gait movements through hand manipulation, and have implemented a mechanism to effectively associate hand movements with overall body movements. We have introduced a network capable of extracting intensity and phase information for each motion frequency, which has demonstrated a certain level of effectiveness. Additionally, we have developed a method to generate gestural motions driven by speech, utilizing pattern matching of their features. This approach enables us to achieve synchronization with speech and diversity in generated motions via pattern-matching at a coarse resolution, while preserving the features of resource samples at a finer resolution. To minimize kinematic errors, we decomposed the data structure into upper and lower body parts and excluded the effect of voice on lower body movements.

研究分野：機械学習に基づく人物動作の操作と制御

キーワード：キャラクタ・アニメーション 音声駆動型身振り制御 身体動作の3次元操作 多重解像度パッチ照合 特徴空間の統合

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

## 様式 C - 19、F - 19 - 1 (共通)

### 1. 研究開始当初の背景

仮想人間を用いた既存の会話システムでは、全身の動きや振る舞いは計測動作をあてはめるか、事前登録された動きをデバイス操作で選択的に再生する方法が用いられている。しかし、前者の方法では計測精度や演技力が品質に大きく影響し、後者の方法では動きのバリエーションが乏しく豊かな表現での振る舞いが困難である。

ヒューマノイド・アニメーション分野で開発された技術で動きを生成する場合、通常の歩行や作業動作に関しては、その移動方向や手先・足先の到達位置を指定すれば直観的に操作できるが、意味を伝える身振りや舞踊等の複雑な動きに対しては、直観的な操作方法は確立されていない。

### 2. 研究の目的

上記の問題を解決するには、入力操作と複雑な身体動作を直観的に関連づける必要があるが、そのためには操作と動作の時系列信号の潜在的な特徴の分布や構造間の整合性が必要となる。本研究では、簡単な操作によって仮想人間の表情豊かな動作や身振りを実時間生成するために、入力信号の効果的かつ直観的な変換機構を開発する。そのために、手指や音声等を用いた対話的な操作に適した特徴空間を学習する機構を考案し、マルチモーダルな入力操作にも適用できる様に、特徴量の分布や構造の異なる複数の入力信号や記号命令に対して、入力と動作の間に意味での一貫性が有る写像を獲得する手法を開拓する。これにより、複雑な身体動作と簡便な操作の潜在変数空間での次元のもつれを無くして、直観的に把握しやすい変換機構を獲得する。そして、表現力が豊かな歩行動作や身振り動作に対して定性的および定量的な評価実験を実施し、その適用可能性を検証する。

### 3. 研究の方法

#### (1) 手指を用いた歩行動作の対話的操作

手指の動きを非接触に計測できるデバイス(図1参照)を入力装置として導入し、歩行動作を様々な歩容で対話的に操作する手法を開発する。光学式のモーションキャプチャ装置を用いて計測した歩行動作のデータセットに対し、それらの動きを直観的に操作するような手指の動きを計測して、対となるデータセットを構築する。また、歩行動作の特徴を自動的に抽出するために、畳み込みニューラルネットワークに基づく符号化器(図2参照)を用いて、関節角度の時系列信号を潜在変数空間に埋め込む。それと同時に、手指の動き情報を異なる符号化器(図3参照)で潜在変数化し、歩行動作と手指の動きが直観的に対応づけられる変換機構を構築する。

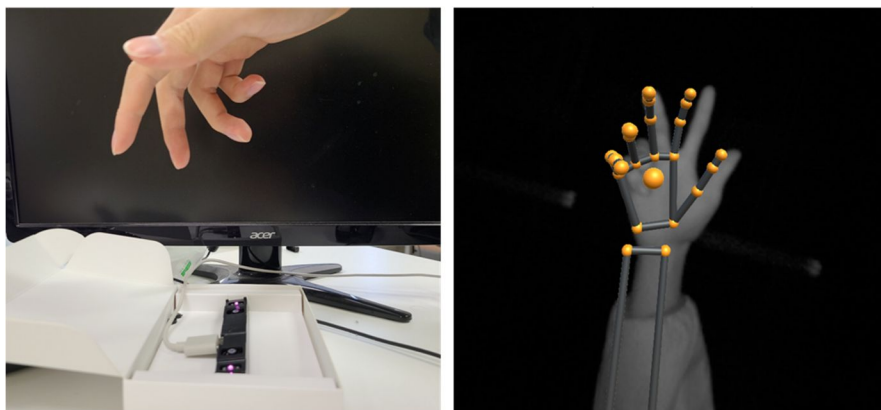


図1 手指の動きを非接触で計測するデバイスと可視化結果

手法の適用範囲を歩行以外の全身動作にも適用するために、動きの周波数毎の強度と位相の情報を抽出できるニューラルネットワーク(文献 )を導入し、その出力を用いて手指の動きと全身動作を柔軟かつ頑健に関連づける機構を開発する。特に、教師なし学習で抽出される位相情報に着目して手指の動きから動作を生成する。すなわち、各種の動作とそれを操作する様に動かした手指のデータから抽出した各周波数での位相変数をパターン照合して(図4参照)変換機構を構築する。さらに、動作生成時の位相チャンネル数の変化の影響を実験的に検証する。

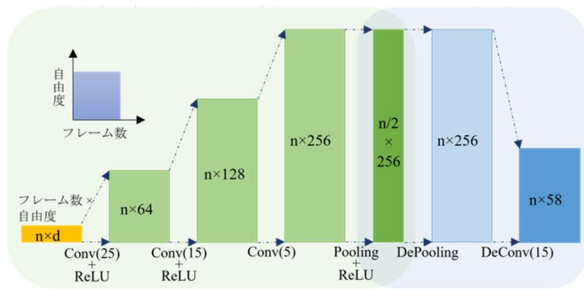


図2 歩行動作に対する符号化器

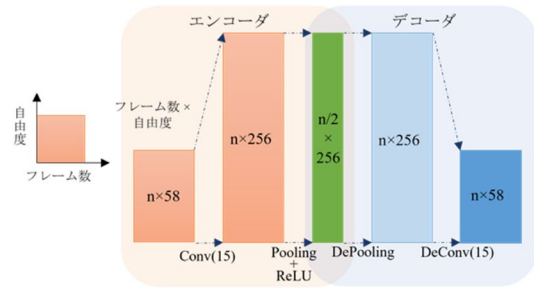


図3 手指動作に対する符号化器

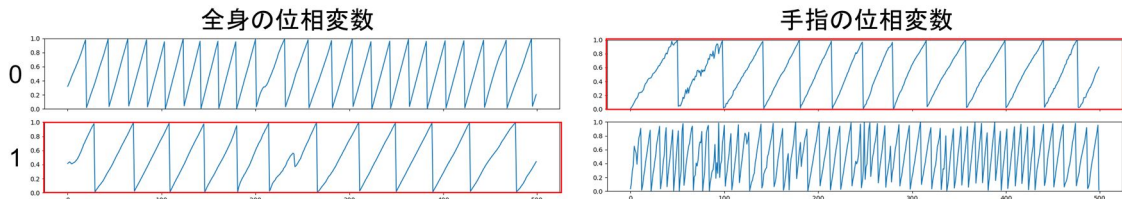


図4 二つの位相チャンネルの出力結果 (赤色枠は照合に用いたチャンネル対)

## (2) 発話を用いた身振り動作の操作

発話音声を手段として身振り動作を操作するために、音声と動作の特徴量の照合により動作生成する手法を開発する。当初の計画では歩行動作の操作と同様に、ニューラルネットワークによる潜在変数への埋め込みを介した変換を開発する予定であった。しかし、区分的な動作データの多重解像度での照合を用いて多様な動作を高品質に自動生成する手法(文献 )が発表され、潜在変数化を介さずに高品質な動作が多様に生成できる可能性が示唆された。したがって、この手法に対して音声データの統合的な照合機構を導入し、音声信号による身振り動作の対話的な操作を実現する。すなわち、データから直接計算される特徴量を多重の時間解像度で再構成し、一定の時間間隔で区切られた動作と音声のデータ(パッチと呼ばれる)間での類似度を、各特徴量で算出される距離の重み和によって求める。そして、正規分布雑音で初期化される動作データのパッチを粗い解像度レベルから詳細な解像度レベルに徐々に逐次置換していくことにより、発話音声から多様な身振り動作を操作する手法を開発する。

## 4. 研究成果

### (1) 手指を用いた歩行動作の対話的操作

位相情報を用いた特徴量の対応づけと照合により、部分的に周期性を有する上半身の動きに対しても、手指の動きで実時間操作することが可能となった(図5参照)。しかし、歩行と同時かつ独立に演じられる上半身の動作を片手的手指での操作のみで逐次制御した場合、部分的に不自然な動きを生じた。また、少ない位相チャンネル数で学習した場合は手指の動きに合致した歩行動作を生成できたが、その動きの自由度は小さく、変化も乏しいものとなった。

一方、全ての位相チャンネルを用いた場合には生成される動作は安定性を欠くものとなり、手指の動きも十分には反映されなかった。これは、全身動作と手指の動きの位相波形の類似度が低かったのが原因と考えられる。すなわち、複雑かつ複合的な動作の制御には位相変数のチャンネル数を多く設定することが考えられるが、その反動として操作が難しくなるので、それらのバランスを深慮して設計する必要がある。

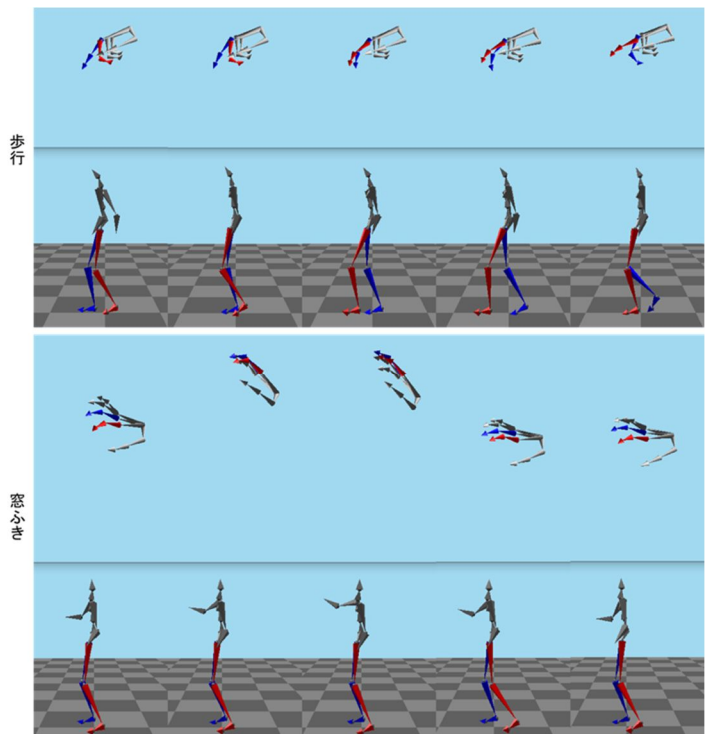


図5 歩行と窓拭き動作の操作結果

(2) 発話を用いた身振り動作の操作

音声と動作の各距離に対する重みを解像度のレベル毎に変化させる(図6参照)ことにより、音声と同期した自然で多様な動作を生成する機構を開発した。初期の粗い解像度のレベルでは、音声間の距離を重視すると身振り動作との同期が保たれて動きの再現性が向上した(図7参照)。一方で、初期状態から音声の影響を除外するレベル数を増加させると、生成動作の多様性が向上した(図8参照)。また、後段の複数の詳細な解像度レベルでは動作間の距離のみを反映させることで、自然で滑らかな動きが得られ、素材として用いたデータの本来の特徴は、重みの設定法に依らずに保持されることが判明した。

生成動作において足先の接地点が若干滑るという欠点が発見されたが、この問題に対しては身体骨格データを上半身と下半身に分離し、パッチ間の照合の際に用いる距離計算の重みを下半身では音声に対する影響を排除することにより解決できた(図9参照)。これらの研究成果に対しては、関連する技術分野での国内の学術会議(情報処理学会・コンピュータグラフィックスとビジュアル情報学研究会)で優秀研究発表賞が授与された。

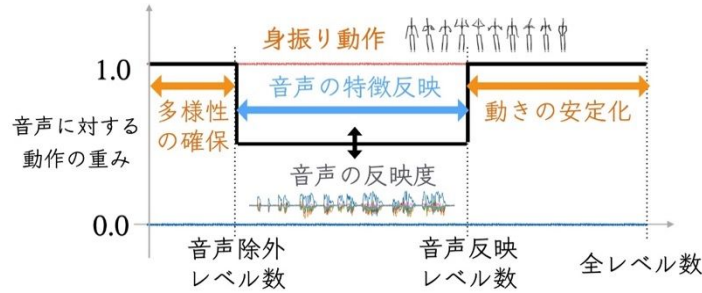


図6 動作と音声の各特徴空間で計算された距離に対する重み係数の設定法

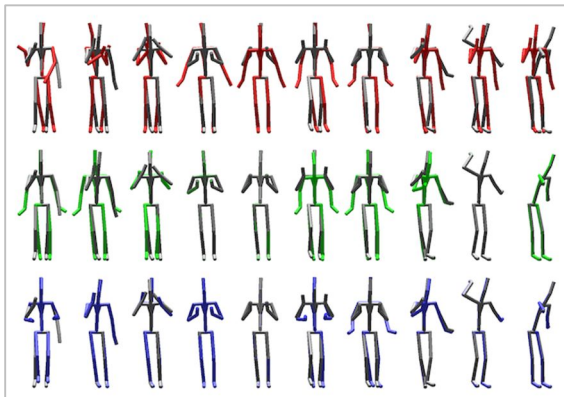


図7 音声を反映する解像度レベル数による再現性の変化(上から順に0, 3, 6)

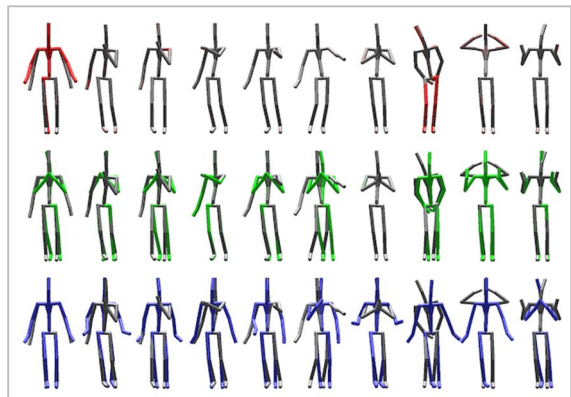


図8 音声を除外するレベル数による多様性の変化(上から順に4, 7, 10)

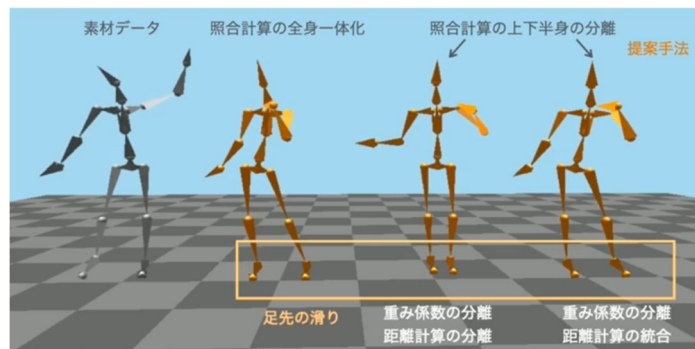


図9 . 重み係数の上下半身分離による足滑り現象への影響

<引用文献>

Sebastian Starke, Ian Mason, and Taku Komura, "DeepPhase: periodic autoencoders for learning motion phase manifolds", *ACM Trans. Graph.* 41, 4, Article 136, pp.1-13, 2022.

Weiyu Li, Xuelin Chen, Peizhuo Li, Olga Sorkine-Hornung, and Baoquan Chen, "Example-based Motion Synthesis via Generative Motion Matching", *ACM Trans. Graph.* 42, 4, Article 94, pp.1-12, 2023.

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2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 Ryusei Nishikawa, Chunzhi Gu, Haruka Takahashi, Shigeru Kuriyama	4. 巻 4
2. 論文標題 Separable Facial Image Editing via Warping with Displacement Vectors	5. 発行年 2024年
3. 雑誌名 IEEE Access	6. 最初と最後の頁 1-11
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/ACCESS.2024.3411699	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 Tomohiko Mukai	4. 巻 11 (3)
2. 論文標題 Transformation Constraints Using Approximate Spherical Regression	5. 発行年 2022年
3. 雑誌名 Journal of Computer Graphics Techniques (JCGT)	6. 最初と最後の頁 1-21
掲載論文のDOI（デジタルオブジェクト識別子） なし	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 -

〔学会発表〕 計11件（うち招待講演 1件 / うち国際学会 2件）

1. 発表者名 栗山 繁
2. 発表標題 音声駆動型モーションマッチングを用いた身振り動作生成
3. 学会等名 情報処理学会コンピュータグラフィックスとビジュアル情報学研究会第193回研究発表会
4. 発表年 2024年

1. 発表者名 Yuhao Dou and Tomohiko Mukai
2. 発表標題 Facial Animation Retargeting by Unsupervised Learning of Graph Convolutional Networks
3. 学会等名 Nicograph International 2024（国際学会）
4. 発表年 2024年

1. 発表者名 西川隆盛、高橋遼、顧淳祉、栗山繁
2. 発表標題 3次元キーポイントのスイープ操作に基づいた顔画像編集
3. 学会等名 情報処理学会第86回全国大会
4. 発表年 2024年

1. 発表者名 Yuhao Dou and Tomohiko Mukai
2. 発表標題 Human Facial Animation Retargeting by Unsupervised Neural Network
3. 学会等名 情報処理学会コンピュータグラフィックスとビジュアル情報学研究会第193回研究発表会
4. 発表年 2024年

1. 発表者名 小南悠人、向井智彦
2. 発表標題 音声データからの感情推定を用いたリップシンクモーションの自動生成
3. 学会等名 Visual Computing 2022 (Poster)
4. 発表年 2022年

1. 発表者名 竇宇豪、向井智彦
2. 発表標題 深層学習による顔アニメーションリターゲティングの検討
3. 学会等名 Visual Computing 2022 (Poster)
4. 発表年 2022年

1. 発表者名 胡 博文、安藤 大地、向井 智彦
2. 発表標題 Analysis and Synthesis of Eye Movements for Conversational Agent
3. 学会等名 第84回情報処理学会全国大会
4. 発表年 2022年

1. 発表者名 瀧川 真太郎、栗山 繁
2. 発表標題 手指の動きを用いたアニメーションの生成
3. 学会等名 Visual Computing 2021 (Poster)
4. 発表年 2021年

1. 発表者名 西川 隆盛、栗山 繁
2. 発表標題 顔画像のスワイプ操作による動画生成
3. 学会等名 情報処理学会コンピュータグラフィックスとビジュアル情報学研究会第185回研究発表会
4. 発表年 2022年

1. 発表者名 Tomohiko Mukai
2. 発表標題 Geometric Deformation of Virtual Character Animation Using Cascaded Spacetime Constraints
3. 学会等名 Mini-symposium on Sensor Network Localization and Dynamical Distance Geometry (Thematic Program on Geometric Constraint Systems, Framework Rigidity, and Distance Geometry) (招待講演) (国際学会)
4. 発表年 2021年

1. 発表者名 向井 智彦、岩本 尚也
2. 発表標題 二段階スキニング分解による頂点アニメーションの近似
3. 学会等名 Visual Computing 2021
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

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	向井 智彦  (Mukai Tomohiko)  (10432296)	東京都立大学・システムデザイン研究科・准教授    (22604)	

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