

令和 6 年 6 月 17 日現在

機関番号：13904
研究種目：基盤研究(B)（一般）
研究期間：2019～2023
課題番号：19H04231
研究課題名（和文）アバターを介した非言語コミュニケーションのための動作スタイルの学習と即時変換

研究課題名（英文）Learning and realtime conversion of motion styles for non-verbal communication via avatar

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） 13,100,000円

研究成果の概要（和文）：ニューラルネットワークによる動作のスタイル転移手法として、白色化を介した潜在変数を一定時間毎に区切った動作パッチのパターン類似度に基づく置換方法が最高性能であることを発見した。また、身振り動作のスタイル学習に用いるデータセットを体系的に構築して一般公開した。その拡張として、スタイル動作のパッチを時間伸縮させながら幅広く参照できるパターン照合手法を開発し、タイミングが異なり変化の激しい動作に対するスタイル転移の性能を向上させた。さらに、トランスフォーマモデルを用いて、トークン化された動作パッチ特徴の前後関係から文脈に相当する情報を抽出し、身振り動作の意味的構造を捉えたスタイル転移手法を開発した。

研究成果の学術的意義や社会的意義
深層学習を用いた動作のスタイル変換は様々な手法が提案されているが、身振りを対象とした手法は発話内容から動作生成する手法の拡張機能として開発された生成的な手法のみであり、その表現力の欠如に問題があった。本研究では、素材となるスタイル動作を潜在変数空間で置換するアプローチにより、生成的手法では成し得なかった高性能なスタイル転移を可能とした点に学術的な意義がある。本研究で開発した技術により身振りの表現力が高められ、新たなデータに対する再学習が不要で生成時の計算量も少ない利点は、仮想的な没入空間における分身（アバター）を用いたコミュニケーション環境の高度化に資するものであり、社会的な意義を有する。

研究成果の概要（英文）：We discovered that the most effective neural network-based motion style transfer method involves replacing patterns of similar motion patches separated by latent variables at regular time intervals after applying whitening. As an expansion, we developed a pattern-matching approach that can handle a large number of style gesture patches with time stretching and shrinking, resulting in improved performance in transferring style for dynamic movements of varying timings. Moreover, we developed a style transfer method that captures the semantic structure of gestural behavior using a transformer model to extract contextual information for tokenized patches. Additionally, we systematically created a publicly available dataset for learning gesture styles.

研究分野：深層学習に基づくヒューマノイド・アニメーション

キーワード：モーションデータ 身振り動作 スタイル転移 ビジョン・トランスフォーマ ヒューマノイド・アニメーション パターン照合 深層学習 動作パッチ

様式 C - 19、F - 19 - 1 (共通)

1. 研究開始当初の背景

V-Tuber 等で利用が拡大している 3DCG アバターの非言語的なコミュニケーションの品質を向上させるためには、外見から受ける印象や感情に即した様態に身振り動作を演出する必要がある。しかし、既存のアニメーション技術は歩行や格闘などの動力的な最適化が可能な動きに適用範囲が限定されており、様態の微妙な違いをスタイルとして捉えて演出できる、実時間での生成・変換技術は未だ実現されていない。また、キャラクタ・アニメーションの制作工程では、モーションキャプチャ装置で計測した動作をそのままアバターに用いているが、動きに外見や感情等に沿った特徴を付与するには、演者に高度な演技力が要求される。また、体型が不自然に誇張された人物や擬人化された動物等、非実在的な体躯のアバターを用いる際には、動作の高度な変換が必要となる。これらの要求を満たすために、動作データ群からスタイル特徴を学習し、動作のスタイルを即時変換する機構の開発が望まれている。

2. 研究の目的

本研究では、動作のスタイルを感性的特徴や外見の印象に関連づけて体系化してモーションキャプチャデータを収集し、それを用いてアバターの個性や表情を生み出す身振り動作の特徴を学習して即時転移する手法を開発する。この技術により仮想身体を用いた社会交流を豊かなものとし、演芸等の新たな形態でのエンタテインメント環境を創出する。

3. 研究の方法

(1) ニューラルネットワークを用いて身振り動作のスタイルを転移させるために、画像の描画スタイルを転移させる再学習を必要としない既存手法(文献)の、身振り動作データに対する有効性を調査する。また、実時間でのスタイル変換を可能とするために、動作の時系列での連続性を考慮した予測的な変換手法を考案する。

(2) タイミングにズレが生じている様なスタイルの身振り動作にも対応できる様に、時間的な伸縮を取り入れたパターン照合の手法を取り入れる。

(3) 動的な変化が激しく様な誇張された身振り動作にも対応できる様に、分割された動作間のパターン照合を、その選択回数の偏りが生じない様に拡張する。

(4) 身振り動作の意味的な構造や前後関係も反映できるニューラル・ネットワークのモデル(文献)を取り入れ、スタイル転移の定量的な評価指標を設計する。

4. 研究成果

(1) 畳み込みニューラルネットワークで潜在変数化された動作データに対して白色化を適用し、それを一定時間毎に区切ってパッチとして扱い、通常とスタイル付けされたパッチデータ間で類似度を計算してパッチ同士を置換することによりスタイル転移する機構(図1参照)を提案した。

本手法では、身振り動作のスタイル変換を画像のスタイル変換で提案されたゼロショットの学習法(文献)を導入し、身体の部位毎に潜在変数化したデータの白色化・彩色化と注意機構に基づくパッチ合成の有効性を示した。また、前時刻のパッチに対して求めた注意機構の重みを、次時刻のパッチに対する重みとして用いることにより、実時間でのスタイル変換機構を実現した。

本手法により、白色化・彩色化とパッチ単位の置換の機構を組み合わせることで、より自然な動きで安定にスタイル転移できることが示された(図2参照)。また、動作データを身体の部位毎に分割して潜在変数化することで、スタイル転移の前後で動きの意味的な特徴を保持する性能が向上することを示した(図3参照)。さらに、パッチ置換の際の重みを先読みする予測的なスタイル変換では、注意機構の導入によって予測誤差を軽減できることを示した(図4参照)。以上の研究成果に対しては、関連する技術分野での国内では最大規模の学会会議(Visual Computing 2020)において、優秀な発表に対して授与される CGVI 賞に選ばれた。

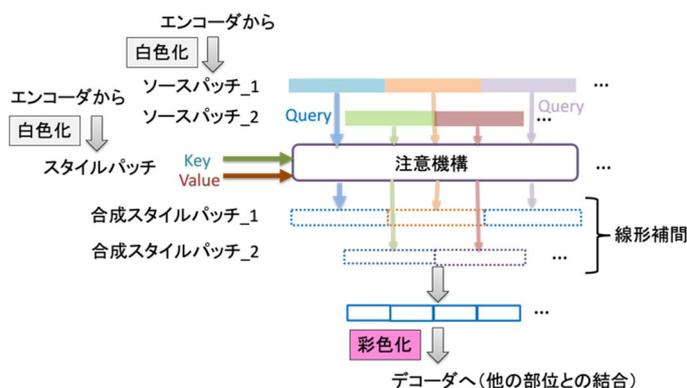


図1 パッチの置換に基づくスタイル転移の概略

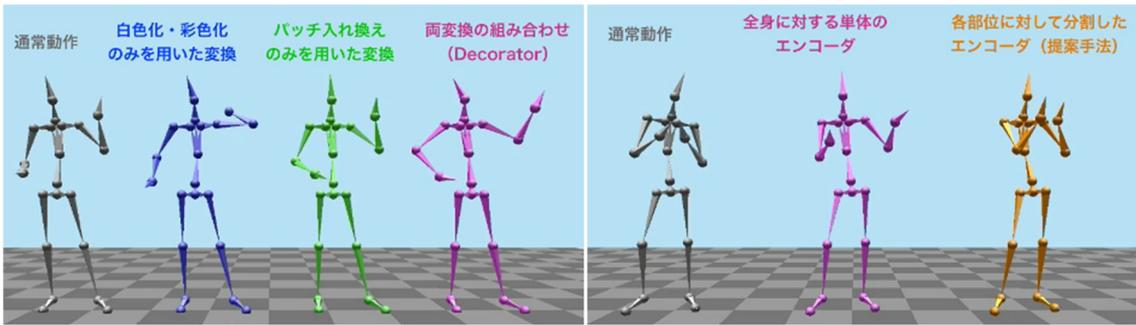


図2 白色・彩色化とパッチ置換のみとの比較 図3 全身と部位毎の潜在変数化の比較

社会的な貢献としては、OCEAN モデルから導出された性格認識の分類基準(文献)に基づいて身振りの動作スタイルを通常、外向的、内向的、安定的、不安定の5種類に分類し、各分類でのスタイル動作を計画的に収集して、研究に利用できるデータアーカイブを構築した。

(2) 上記(1)の手法ではスタイルの時間軸に沿った特徴を転移できないので、動作のタイミング特徴を転移する際に、伸縮可能なパッチを導入した(図5参照)。この拡張手法により、タイミングの異なるスタイル動作に対しても、以前の手法よりもスタイルが高品質に転移されることを示した(図6参照)。

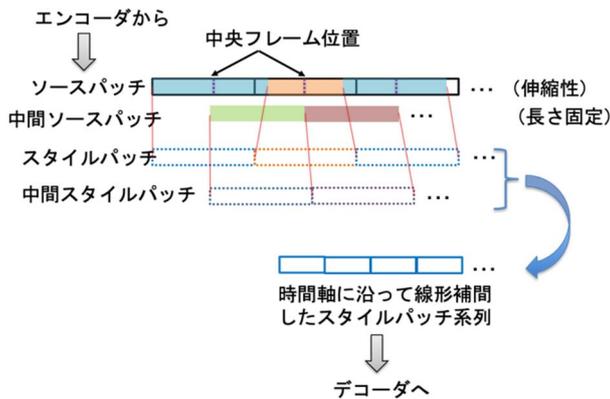


図5 伸縮性のパッチを用いた潜在変数の置換

(3) 上記(1)の手法では、スタイル転移前の身振り動作の意味内容を保持するためには、置換するためのスタイル動作のパッチにも内容的に十分なバリエーションが必要とされる。この問題を解決するために、階層的なオートエンコーダを用いたパッチの置換機構を導入して変換に柔軟性を持たせ、さらにパッチの使用頻度を平準化できる再配置アルゴリズムにより、表現としての大域的な構造をも転移できる変換機構を構築した。

本研究では、エンコーダで計算された特徴マップを多段階でデコーダに反映させるために、エンコーダの各層で出力される特徴量をデコーダの対応する層の特徴量と連結するスキップ接続を有する U-Net を採用した(図7参照)。この多重解像度の変換により潜在変数の時間軸での縮退に不変な照合計算ができ、変動の激しい動作間の対応付けも可能となった(図8参照)。また、スタイル転移の際に特定のパッチのみが頻繁に選択されてしまう問題に対しては、使用頻度の偏りを解消しながら類似度も併せて考慮できる平準化アルゴリズムを開発した(図9参照)。

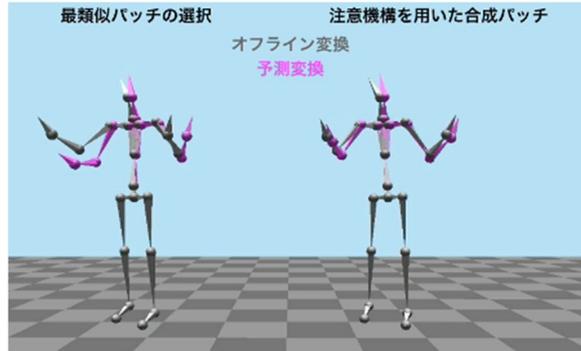
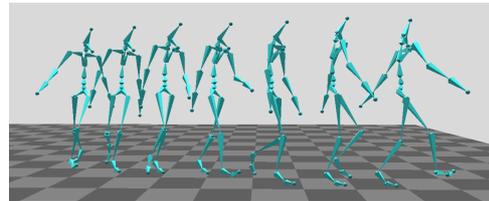
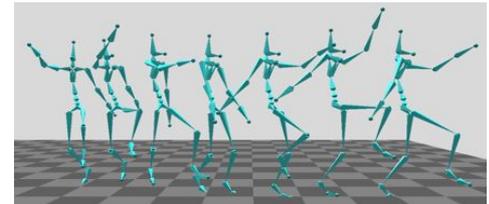


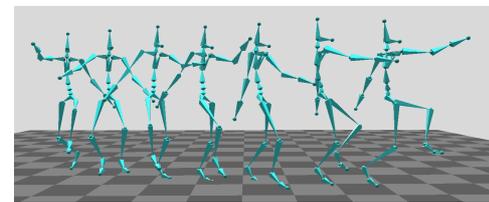
図4 注意機構を用いた予測変換の効果



(a) 転移前の回旋歩行



(b) 伸縮性の無い固定パッチでの転移



(c) 伸縮性パッチでの転移

図6 伸縮性パッチを用いたスタイル転移

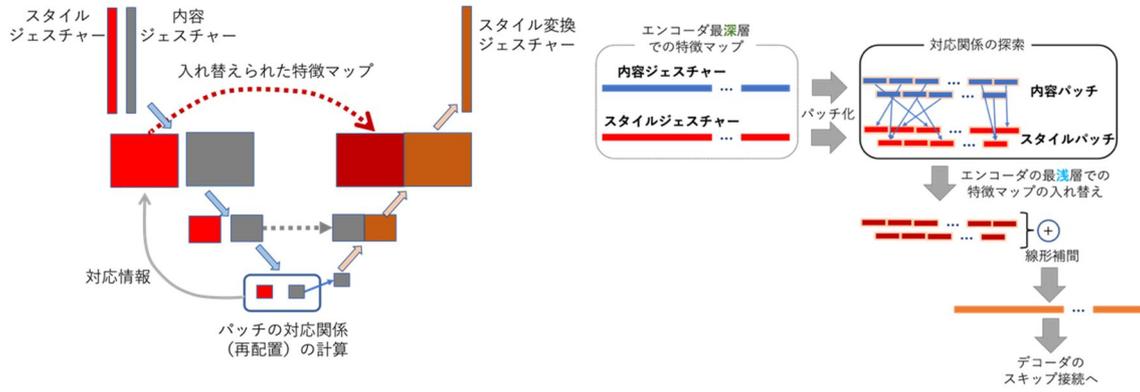


図7 U-Net 型のアーキテクチャを用いたオートエンコーダの構成とスタイル変換の流れ

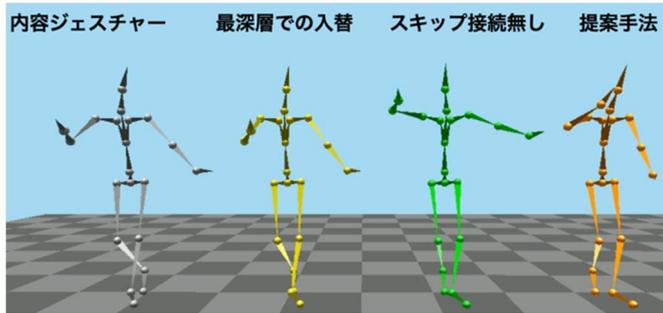


図8 スキップ接続を用いた多重解像度変換の有効性

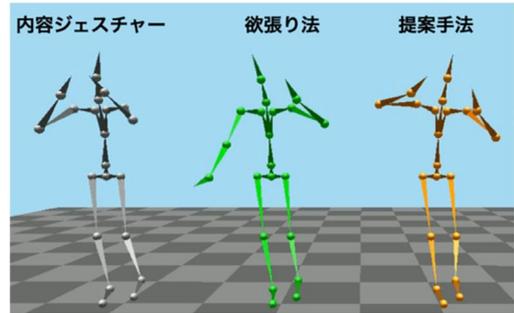
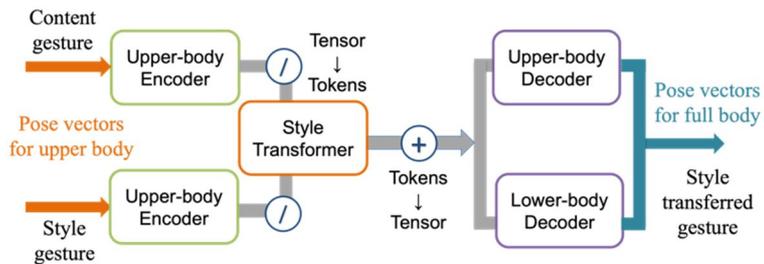
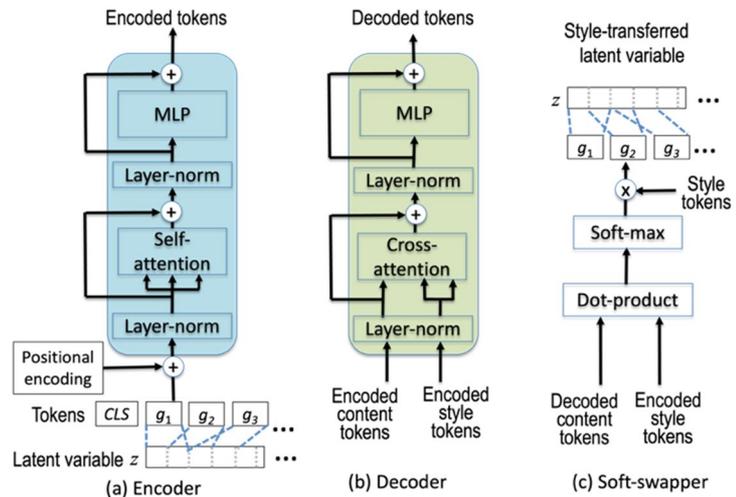


図9 スタイルパッチの使用回数平準化

(4) 上記の手法のさらなる拡張として、パッチ化された身振り動作を単語(トークン)として捉えて動作トークン間の前後関係から文脈に相当する情報をビジョントランスフォーマモデル(文献)により抽出し、身振り動作の意味的構造を捉えたスタイル転移手法(図10参照)を開発した。ここでは、動作の大域的な特徴量を獲得できるクラストークンによるスタイルの非類似度と注意機構内部の潜在変数の自己相似量を用いた構造の非類似度を損失関数として学習することにより、既存手法に対する優位性が得られた(図11参照)。この研究成果は、世界的に著名な学会(Symposium on Computer Animation 2022)で口頭発表し、関連分野における国内最大規模の学会(Visual Computing 2022)においても招待講演として口頭発表した。



(a) 全体の流れ



(b) Style Transformer 部の構成

図10 ビジョン・トランスフォーマによるスタイル転移

<引用文献>

L. Sheng, Z. Lin, J. Shao and X. Wang, "Avatar-Net: Multi-scale Zero-Shot Style Transfer by Feature Decoration," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp.8242-8250.

Harrison Jesse Smith and Michael Neff, "Understanding the impact of animated gesture performance on personality perceptions," *ACM Trans. Graph.* 36, 4, Article 49, 2017, pp.1-12.

N. Tumanyan, O. Bar-Tal, S. Bagon and T. Dekel, "Splicing ViT Features for Semantic Appearance Transfer," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp.10738-10747.

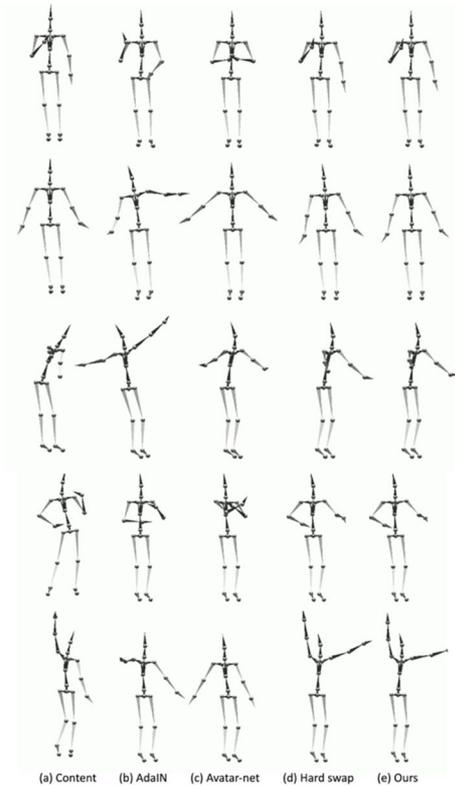


図 1 1 既存手法との生成結果比較

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 1件）

1. 著者名 Kuriyama Shigeru, Mukai Tomohiko, Taketomi Takafumi, Mukasa Tomoyuki	4. 巻 41
2. 論文標題 Context based style transfer of tokenized gestures	5. 発行年 2022年
3. 雑誌名 Computer Graphics Forum	6. 最初と最後の頁 305 ~ 315
掲載論文のDOI (デジタルオブジェクト識別子) 10.1111/cgf.14645	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 -

1. 著者名 Gu Chunzhi, Zhang Chao, Kuriyama Shigeru	4. 巻 150
2. 論文標題 Orientation-aware leg movement learning for action-driven human motion prediction	5. 発行年 2024年
3. 雑誌名 Pattern Recognition	6. 最初と最後の頁 110317 ~ 110317
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.patcog.2024.110317	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17件（うち招待講演 1件 / うち国際学会 2件）

1. 発表者名 Kuriyama Shigeru, Mukai Tomohiko, Taketomi Takafumi, Mukasa Tomoyuki
2. 発表標題 Context based style transfer of tokenized gestures
3. 学会等名 Symposium on Computer Animation 2022 (国際学会)
4. 発表年 2022年

1. 発表者名 栗山 繁、向井 智彦、武富 隆文、武笠 知幸
2. 発表標題 身振りの文脈を考慮した動作のスタイル変換
3. 学会等名 Visual Computing 2022 (招待講演)
4. 発表年 2022年

1. 発表者名 若生 翼、栗山 繁、高橋 遼
2. 発表標題 化粧工程の分解に基づく顔画像の化粧スタイル転写
3. 学会等名 Visual Computing 2022
4. 発表年 2022年

1. 発表者名 周山哲也、栗山繁、高橋遼
2. 発表標題 注意機構を用いた風景画像の浮世絵風スタイル変換
3. 学会等名 電子情報通信学会総合大会
4. 発表年 2023年

1. 発表者名 若生翼、高橋遼、栗山繁
2. 発表標題 化粧工程の分解に基づく顔画像の化粧スタイル転写
3. 学会等名 情報処理学会第85回全国大会
4. 発表年 2023年

1. 発表者名 刀根 千尋、高橋 遼、栗山 繁
2. 発表標題 絵画に対する感情表現を考慮したスタイル変換
3. 学会等名 第185回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
4. 発表年 2022年

1. 発表者名 栗山 繁
2. 発表標題 パッチの再配置に基づくジェスチャーのスタイル変換
3. 学会等名 Visual Computing 2021
4. 発表年 2021年

1. 発表者名 栗山 繁
2. 発表標題 注意機構を用いた身振り動作のスタイル変換
3. 学会等名 Visual Computing 2020
4. 発表年 2020年

1. 発表者名 向井智彦
2. 発表標題 3次元SRT変換の例示ベース連動制御
3. 学会等名 Visual Computing 2020
4. 発表年 2020年

1. 発表者名 栗山 繁
2. 発表標題 伸縮性パッチを用いた動作スタイルのゼロショット変換
3. 学会等名 第177回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
4. 発表年 2020年

1. 発表者名 幸家 大和、遠藤 結城、栗山 繁
2. 発表標題 UkiyoGAN: 自然画像の浮世絵風スタイル変換
3. 学会等名 第177回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
4. 発表年 2020年

1. 発表者名 Mukai Tomohiko, Kuriyama Shigeru, Oshita Masaki
2. 発表標題 Motion Adaptation with Cascaded Inequality Tasks
3. 学会等名 MIG 2019: ACM Conference on Motion, Interaction, and Games (国際学会)
4. 発表年 2019年

1. 発表者名 出口風人、遠藤結城、栗山繁
2. 発表標題 RNNを用いたデータ駆動型モーションリターゲット
3. 学会等名 第18回情報科学技術フォーラム
4. 発表年 2019年

1. 発表者名 木佐省吾、遠藤結城、栗山繁
2. 発表標題 人の関節回転角系列からの身体中心位置の推定
3. 学会等名 第18回情報科学技術フォーラム
4. 発表年 2019年

1. 発表者名 木佐省吾、遠藤結城、栗山繁
2. 発表標題 人の姿勢系列情報からの身体中心位置の推定
3. 学会等名 Visual Computing 2019
4. 発表年 2019年

1. 発表者名 鬼塚慎吾、遠藤結城、栗山繁
2. 発表標題 流れ場に基づくワーピングによる深層生成モデルを用いた人物動画生成
3. 学会等名 第18回情報科学技術フォーラム
4. 発表年 2019年

1. 発表者名 鬼塚慎吾、遠藤結城、栗山繁
2. 発表標題 流れ場生成モデルを用いたワーピングによる人物動画生成
3. 学会等名 Visual Computing 2019
4. 発表年 2019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)	

6. 研究組織（つづき）

	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	遠藤 結城 (Endo Yuuki) (00790396)	筑波大学・システム情報系・助教 (12102)	

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