伸縮性パッチを用いたモーションスタイルのゼロショット変換

栗山 繁

1. はじめに

キャラクタ・アニメーションにおいて自然な動きを生成するのにモーションキャプチャデータが広く用いられるが，その表現力を高めるために動きの微細な特徴（以後，スタイルと呼ぶ）を変換する手法が提案されている．

動きのスタイル特徴を線形システムで同定する手法[1]や周波数領域で最適化する手法[2]等は，スタイル変換するモーションデータ間で動きの特徴的なタイミングをタイムワーピング等で調整して合わせておく必要がある．また，周期的な歩行動作などには適用できるが，歩行と他の動作を組み合わせた様な複合的な動きには柔軟に対応できない．

近年提案された深層学習に基づく動作スタイルの変換手法は，画像における描画スタイルの変換手法をモーションデータに適用しており，潜在変数の統計量をグラム行列で変換させた手法[3]や，敵対的生成型ネットワークを導入した手法[4]が提案されている．これらの手法は，画像のスタイル変換がモーションデータにも適用できることを明らかにしたが，スタイル毎の学習が必要な点が問題点として指摘される．

その後，画像のスタイル変換においては，スタイル毎の学習が不要なゼロショット学習に相当する手法が提案されている．統計量に基づく手法の拡張としては，平均と分散共分散の正規化に基づく白色化処理とその逆変換である彩色化処理に基づく手法[5]が提案され，多段階でのスタイル調整が可能な高品質な変換を実現している．また，潜在変数の特徴が類似した一定サイズの矩形領域（パッチ領域と呼ばれる）毎の入れ替えを用いたゼロショット学習に基づく自然なスタイル変換[6]が提案されている．さらに，これらの白色・彩色化処理とパッチ毎の特徴入れ替えの処理を組み合わせた手法[7]も提案されている．本提案手法では，この組み合わせ手法[7]をモーションキャプチャデータのスタイル変換に適用し，既存手法では不可能なスタイルに応じた時間の伸縮を扱えるように拡張する．

2. スタイル変換手法の概要

本研究においては，モーションキャプチャデータ（以後，単にデータと呼ぶ）のスタイル変換は，図１に示す様に，データの潜在変数化を行うネットワークとしてのオート・エンコーダと，それによって潜在変数をスタイル変換するスタイル・コンバータ（図１の特徴変換部）から変換機構を構成する．

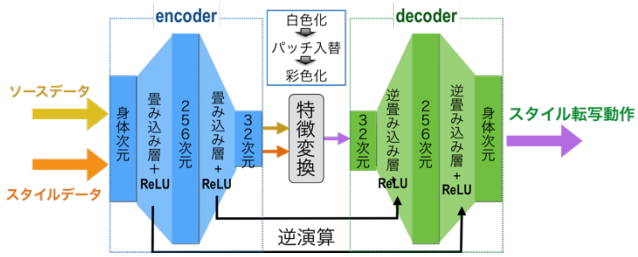


図 1　スタイル変換機構の構成．

2.1 オート・エンコーダ

今回用いたオート･エンコーダはエンコーダ部とデコーダ部から構成され，各々２層の１次元畳み込み（カーネルのサイズは 15 と 1）と ReLU 活性化関数から構成される（図 1 参照）．ここで，デコーダの１次元逆畳み込み層の重み値とバイアス値はエンコーダの畳み込み層と共有される（ただし，逆演算となる様にその値の転置と負数が用いられる）．チャネル数は姿勢次元数を 256 次元に拡張した後に 32 次元に縮退させて潜在変数を算出する．このネットワークは，以下に述べるスタイル・コンバータとは独立に訓練される点に注意されたい．

2.2 スタイル・コンバータ

データに動作のスタイルを付与するには，その対象となる動作データ（以後，ソースデータと呼ぶ）と，それにスタイルを付与する特徴を有する動作データ（以後，スタイルデータと呼ぶ）に対し，オート・エンコーダのエンコーダ部によって潜在変数を個別に算出する．ただし，少ないデータからでも頑健にスタイル特徴を抽出するために，データに対して鏡面対称の動きに変換した値も加えて潜在変数を求める．

ソースデータとスタイルデータに対して算出された潜在変数は，このスタイル・コンバータにより白色化された後に，一定のフレーム間隔毎のパッチに分割される．その後，ソースデータの各パッチでの特徴が最も類似したスタイルデータのパッチを探索し，それらを入れ替える処理を施す．さらに，入れ替えられたパッチを結合して得られる潜在変数をスタイルデータから抽出された特徴で彩色化し，オート・エンコーダのデコーダ部で動作データに変換する．

2.3 データ表現

深層学習を用いたモーションキャプチャデータの既存の変換手法では，入力データとして身体の中心関節からの相対的な３次元位置を対象としたものが多い. しかしながら，関節位置を独立した値として扱うと身体の部位の収縮が生じてしまうので，長さを保持するための調整機構をネットワークに組み込む必要がある，これには入力データで用いた骨格を標準的な身体に正規化して関節位置を計算する前後処理が必要とされ，身体の各部位の比率の差異には対応できない．また，原理的にはネットワークの出力だけで伸縮の誤差を完全に解消できる保証は得られず後処理が必要となり, 生成結果のスタイルに悪影響が及ぶ可能性がある．

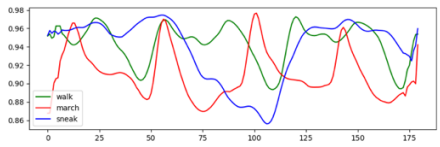
ゆえに本研究では，関節の３次元位置を用いる代わりに，関節の回転角を用いる．回転角の表現で一般的に用いられている四元数は，距離空間の歪みが少なく特異点における不連続性を回避できる等の利点があるが，回転を表すには４次元ベクトルのノルムを１に調整しなければならないので，ネットワーク内にこの正規化処理を加える必要がある[8]．そこで今回は，四元数の対数写像によって得られる指数マップを用いる．これは，回転軸を表す単位３次元ベクトルに回転量をスカラー倍した値に相当する．

本手法では, 中心関節の位置と向きも変換する必要があるが，これらはスタイル特徴として変換すると不自然な動きとなるので変換の対象外とし，地面と水平方向の移動ベクトルはソースデータから，鉛直方向の値はスタイルデータからコピーして利用する．

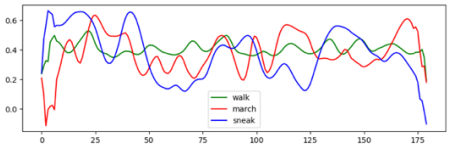
2.4 パッチの伸縮化の導入

動作スタイルの変化に伴いその速度やタイミングも変化するが，既存の手法においてはソースデータのタイミングと同じになるようにスタイルモーションの特徴が時間伸縮（タイムワープ）されて変換される．これは，白色化・彩色化処理においては統計的な特徴のみが変換されるので時間軸に沿った値には影響を与えないのと，特徴の入れ替えにおいてもソースデータの特徴に最も合致するスタイルデータのパッチが選択されるので，その時間的なタイミングもソースデータの特徴と類似したパッチと入れ替えられる事が原因と考えられる．

通常の歩行, 行進および忍び足風のスタイルが付与された歩行の時間軸に沿った 3 秒間の特徴変化を，それらの潜在変数の値と白色化後の値において，時間軸での平均値とのコサイン類似度として求めた結果を図 2 に示す. 図 2 (a) の結果より, 各歩行スタイルで周期が異なる事が確認できる. さらに図 2 (b)の結果より，白色化した値は動作特徴が，ドメイン（歩行スタイル）の違いによる値の偏りが少なく，より複雑なパターンで抽出されている事が確認できる．



(a)　各歩行スタイルに対する潜在変数値の時間変動



(b)　白色化された潜在変数値の時間変動

図 2　通常歩行、行進、忍び足の，潜在変数空間における平均値とのコサイン距離.

そこで本手法では, 潜在変数の空間でフレーム間隔毎に区分化されたソースデータの各パッチの長さを一定の比率で伸縮させ，最も類似したスタイルデータのパッチを探索する．ただし，スタイルデータの側のパッチは伸縮させずに固定長とする点に注意されたい．この処理は，スタイルデータの特徴を反映したタイミングのデータが最適に選択されるように，ソースデータの時間タイミングを伸縮（タイムワープ）させることに相当する．今回の実験では，各パッチ区間に含まれるフレーム数の 25%を増減させて伸縮の上下限とし，４フレーム毎に区間内のフレーム数を変化させ，時間軸で同じ次元数となる様に再標本化を行った．この伸縮して再構成したソースデータのパッチに対して，総当たりで特徴の最も類似したスタイルデータのパッチを，既存手法[5]と同様にコサイン距離を用いて探索した．

区間ごとに探索したパッチをそのまま接続すると, 区間の切り替わるフレームで動きが不連続になってしまうので，画像での手法と同様に区間の半分が重複するパッチの系列（図 3 の中間ソースパッチ）も探索によって求め，二つの系列を時間軸に沿って線形に補間することによって，デコーダに渡す潜在変数の系列を求める．ただし，中間のソースパッチの長さは伸縮したソースパッチの長さとの整合性を考慮して，隣接する二つの伸縮パッチ中央フレームを開始／終了フレームとする長さに固定する（図 3 参照）.

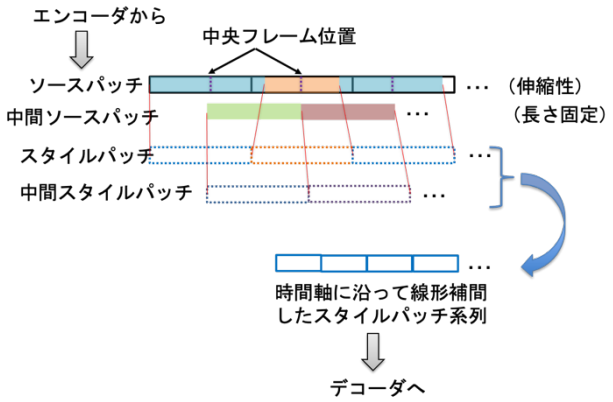


図 3　伸縮性パッチを用いた潜在変数の変換.

実験

今回は，通常の歩容と様々なスタイルの歩容を計測した独自データセット（60 フレーム/秒）を使用した．人体は図 4 に示す 30 関節で構成され，回転角度の計算は☆印の 6 個の末端関節を除く 24 関節に対して計算した．

オート・エンコーダの学習には 35 ファイル（総計 1562 秒分）の様々なスタイルの歩行データを用い，Adam 最適化を用い，バッチサイズは 16，学習率は 10^{-4} に設定した．また，学習用のデータは 240 フレーム分の区間を等確率に標本化した．

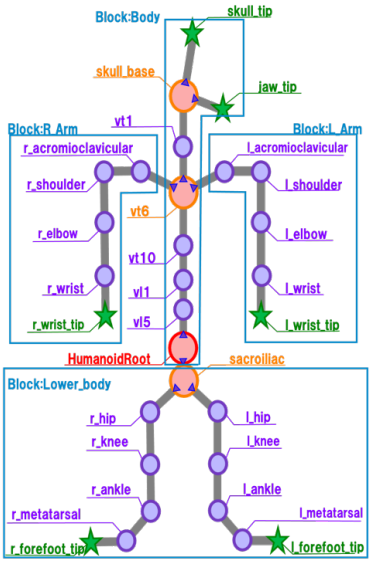


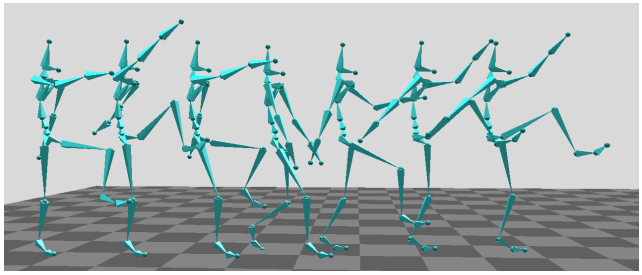
図 4　使用した動作データの関節構成図．

3.1 スタイルの変換結果

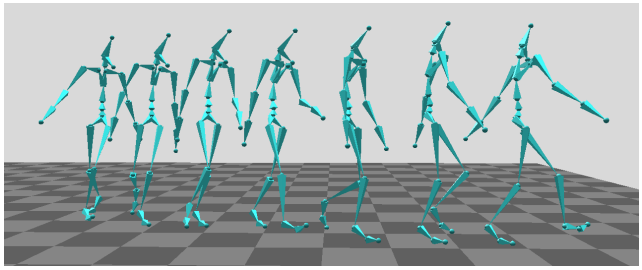
図 5 において，ソースデータとして起動が回旋した歩行を，スタイルデータとして行進調と忍び足の歩行データを用いて実験した結果を示す．ただし，全て 10 フレーム（1/6 秒）毎の標本化であり，図 5 (a) を除く表示は同じフレームの姿勢系列を抜き出したものである．この結果より，図 5 (d) で示すパッチに伸縮性を導入したスタイルの変換結果は，手の振りのタイミングが上段のスタイルデータと一致している点に注意されたい．一方，図 5 (c) で示すパッチに伸縮性の無い既存手法での変換結果は，図 5 (b) のソースデータのタイミングと同期していることが確認できる．

3.2 潜在変数の補間による影響度の調整と遷移

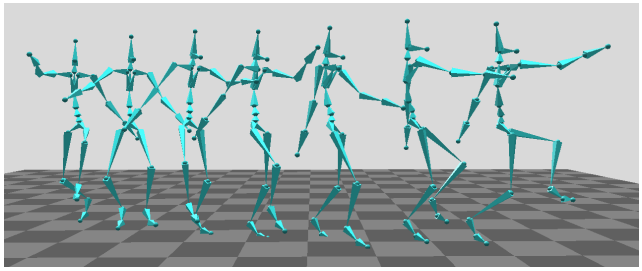
本手法はスタイル特徴を潜在変数空間で変換するものであるが，変換前の値と変換後の値を補間することにより，特徴の反映度合いを調整することが可能である．さらに，潜在変数の値はフレーム毎に計算されるので，経過時間に沿って補間するだけで，特徴が徐々に変化する様な遷移動作が自然に生成される．さらに，スタイルが付与された歩行の潜在変数同士においても上記の様な補間計算が可能であるので，異なるスタイル間の融合や遷移も簡単に計算できる（図 6 参照）．



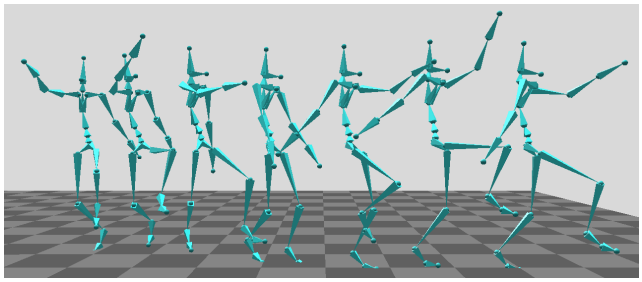
(a)　行進調のスタイルデータ



(b)　回旋歩行のソースデータ

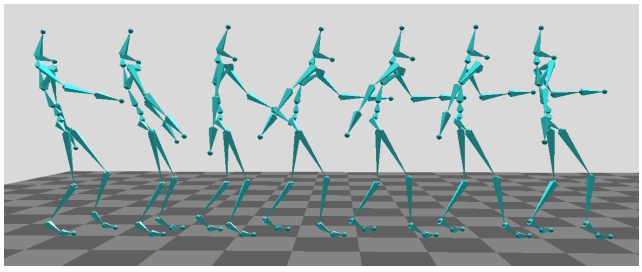


(c) パッチに伸縮性が無い場合の変換結果

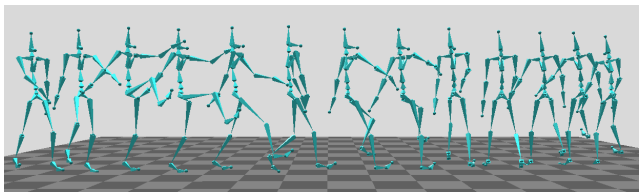


(d)　パッチに伸縮性を導入した変換結果

図 5　スタイル変換の生成例．各姿勢の配置間隔は重なりを避けるために意図的に広げている．



(a)　忍び足のスタイルデータ



(b) 忍び足から行進調（図 3 (b)）への遷移動作

図 6 時間に沿ったスタイル遷移例.

3. おわりに

本報告では, 潜在変数空間での特徴量の入れ替えに伸縮可能なパッチを用いた探索を導入することにより，スタイルの時間的な特徴変化も補足できるスタイル変換手法を提案した．本提案手法では再学習も不要であり，ソースとスタイルのデータから潜在変数を計算して数値処理するだけなので，対話的な即時変換にも適している．しかしながら，中心関節の位置に関しては推定精度が保証されないので，足先が地面と接触しなかったり滑ったりする現象を除去できない．これは，深層学習に基づく既存の手法とも共通する欠点であるが，中心関節の位置推定機構等を新たに導入して，その誤差を解消することが今後の課題である．

伸縮性パッチの探索に関しては, 今回の実験では区間内のフレーム数を 4 フレーム毎に離散的に変化させて再構成したパッチ間で総当たり的に類似特徴を探索したが，この計算機構を 1 フレーム毎に高精度化し，高度な最適化計算を導入して探索計算を効率化する事も重要である．

複合的な動作に対するスタイル特徴の精確な変換は, これまでに前例のない挑戦的な課題であり，動作の種類や関節毎のスタイル変換が必要になる．用いているオート・エンコーダに，関節毎に分解されたグラフ的なネットワーク等の新たな構造を導入する事も，今後の検討課題である．

参考文献

[1] S. Xia, C. Wang, J. Chai, and J. Hodgins, “Realtime style transfer for unlabeled heterogeneous human motion,” *ACM Trans. Graph.*, vol. 34, no. 4, pp. 119:1-119:10, 2015.

[2] M. E. Yumer and N. J. Mitra, “Spectral style transfer for human motion between independent actions,” *ACM Trans. Graph.*, vol. 35, no. 4, pp. 1–8, 2016.

[3] D. Holden, T. Komura, and I. Kusajima, “Fast Neural Style Transfer for Motion Data,” 2017.

[4] 渡祐貴，中澤篤志，幸村琢, MotionGAN: 関節パラメータの敵対的学習による動作スタイル生成, 研究報告コンピュータグラフィックスとビジュアル情報学（CG）, 2019-CG-176(23)

[5] Y. Li, C. Fang, J. Yang, Z. Wang, X. Lu, and M.-H. Yang, “Universal Style Transfer via Feature Transforms,” 2017.

[6] T. Q. Chen and M. Schmidt, “Fast Patch-based Style Transfer of Arbitrary Style,” 2016.

[7] L. Sheng, Z. Lin, J. Shao, and X. Wang, “Avatar-Net: Multi-scale Zero-Shot Style Transfer by Feature Decoration,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 8242–8250, 2018.

[8] D. Pavllo, D. Grangier, and M. Auli, “QuaterNet: A Quaternion-based Recurrent Model for Human Motion,” pp. 1–13, 2018.