Vol.2019-CVIM-217 No.31
2019/5/30

時空間同期の取れた2人協調動作生成

前田 孝泰[1,a)]　浮田 宗伯[1,b)]

概要：近年 VR，AR 技術とヒューマノイドロボットが発展している．しかし，現在の多くのキャラクターやロボットの動きはプログラムされた固定的なものであり，人とロボットがハイタッチや協力して物を運ぶなどの協調動作を行うには，ある程度人がロボットへタイミングと位置を合わせることが必要である．本研究の最終目標は，ユーザに対応する協調動作をリアルタイムに生成することである．これにより, ユーザがヒューマノイドロボットや仮想空間内のキャラクターとより高度な交流を行うことができる．協調動作には，位置やタイミングを考慮した動作を生成することが必要である．これを実現するため，学習データの特徴をとらえた出力が期待できる深層学習技術を利用した．本研究では，2 人のフィストバンプ動作を記録し，学習データとした．これを予測することで，学習データ内の人同士の位置とタイミングを捉えた協調動作の生成ができる．本研究では，入力動作と生成動作の時空間同期を取ることを目指した新たな協調動作生成手法を提案する．この手法は，生成動作の不連続を抑制する残差パス，また関節の位置と関節間距離を同時に学習する新たな損失関数を持った Reccurent Neural Network（RNN）である．さらに，新たに作成したフィストバンプ動作のデータセットで提案手法の有用性を検証した．その結果，従来法と比べ二乗平均平方根誤差を 7.95[cm] から 3.18[cm] へ低減することができ，より時空間同期の取れた協調動作を生成することができた．

1. はじめに

近年 VR 技術やヒューマノイドロボットが発展しており，ヘッドマウントディスプレイを装着しながら楽しむ VR ゲーム, NAO® や Pepper® (SoftBank Robotics 社 [1], [2]) をはじめとする人と触れ合うヒューマノイドロボットが普及している．しかし，これらの VR ゲーム内のキャラクターとヒューマノイドロボットには，人との協調動作を行うことが難しいという課題がある．なぜなら，現在のキャラクターやロボットの動作の多くは，対する人の動作を考慮しない固定的なものであるため，ある程度人がキャラクターやロボットの動きに合わせに行かなければハイタッチや物を渡すなどの協調動作を行えない．今後，ヒューマノイドロボットに介護などといった，より高度な協調動作をさせる場合は，このような固定的な動作ではなく，相手の動作に基づいて相手に対応する動作を生成することが考えられる．これにより，VR ゲームにおけるキャラクターとの高度な交流やヒューマノイドロボットと共に物を運ぶなどの協調作業が実現される可能性がある．よって，本研究の最終目標は，ハイタッチや物を渡すなどの協調動作において，人の動作からその人に対応する動作を生成することである．

1 人の動作の生成は，事前に記録しておいた動作データセットの中から目的とする動作に最も近いものを検索することで実現できる．しかし，2 人以上での協調動作を実現するためには，動作を事前に記録しておくことは現実的ではない．なぜなら，ユーザの動作の多様性を考えると，協調動作は無数に存在してしまうためである．ハイタッチを例に取ると，手を合わせる位置，左右の手どちらを使用するのか，手を上げるタイミング，手を動かす速度などに自由度が存在してしまい，これらすべての多様性に対応できるように大量の動作を記録しておくことは記録するためのコストとデータ量の観点から難しいからである．さらに，もし大量の協調動作を記録できたとしても，目的の動作を大量の動作の中から実時間内に検索しキャラクターやロボットに適応することは難しい．このように，2 人以上が動作する協調動作は単純な動作の再生では実現できない．

また，動作生成を実現するために考えられる手法として，Inverse Kinematics（IK)[3], [4], [5] を用いて目的位置を指定し，そこへ手などを移動させるということも考えられる．IK は，人体などの関節物体を制御する際に，端点の位置を指定しそれに合うように各関節の位置を逆算することで，指定した位置に端点が位置する関節物体の姿勢を計算する

1 豊田工業大学
Toyota Technological Institute
a) sd19445@toyota-ti.ac.jp
b) ukita@toyota-ti.ac.jp

ⓒ 2019 Information Processing Society of Japan

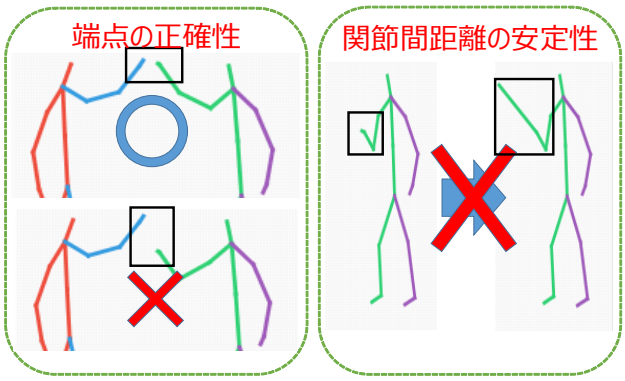

図 1 端点座標の正確性と関節間距離の安定性が求められる理由．左図は端点座標の正確性が求められる理由を示している．左図のようにハイタッチやフィストバンプなどの協調動作を行う際には，黒枠内の端点同士が合うことが求められる．右図は関節間距離の安定性が求められる理由を示している．動作内で黒枠内の関節間距離（腕の長さ）が変化してしまうことは，たとえ端点座標が正確であったとしても人体の構造の観点から不自然である．

技術である．これは，正確に目的の位置に端点を合わせられるという長所があるが，動作を開始する姿勢と目的の位置に端点を置いた姿勢をつなぐ動作に経路，速度やタイミングの自由度が発生してしまうという短所がある．つまりこの手法だけでは，協調動作に必要なタイミングや速度を考慮できない．直線的で等速な動きを生成することはできるが，人を模擬したキャラクターやヒューマノイドロボットの動作生成には適さないと考える．

近年では深層学習が著しく発展してきた．この深層学習では，大量の入力データと正解データのペアを学習することで，入力データと正解データに含まれる関係性や特徴を捉え，新しい入力データを受け取った時に学習データの特徴を考慮した新しい出力データを得ることができるという利点がある．よって，実際の複数人の協調動作を記録したデータセットを作成し，深層学習器に学習させることで，複数人の動作間の関係性や特徴を考慮した動作を生成できる．これにより，データセットに含まれる人同士の位置とタイミングの関係性と，人特有の速度と動作経路を両立した動作を得ることができると考えた．

協調動作には，図 1 に示すように手などの端点座標の正確性と関節間距離の安定性が求められる．端点座標の正確性は，ハイタッチや物を渡す協調動作において端点座標が相手と合うために必要である．また関節間距離の安定性は，人体の構造の観点から隣り合う関節の距離は一定であるため必要である．

本研究の新規性は以下の 2 点である．

(1) 協調動作生成のための新しい深層学習モデルを提案した．また，2 人がフィストバンプを行っているデータセットを作成し，それに対して，生成結果を協調動作生成を行っている従来手法と比較した．

(2) 端点座標の正確性と関節間距離の安定性を両立する誤差評価法を提案した．

2. 関連研究

CG アニメーション，協働ロボット，VR などへの応用が期待され，人体の動作を扱う動作生成と動作予測は CG とロボット工学の分野で盛んに研究されてきた．この節では，この 2 つの技術について関連研究を紹介する．その後，2 人相互動作生成を行った先行研究について詳しく述べる．

2.1 動作生成

動作生成は，検索ベースと学習ベースの 2 種類に大きく分けることができる．

検索ベースの手法 [6], [7], [8] では，生成したい動作に近い動作データを事前に準備したものから探索し，それを IK などで修正することで動作生成を行うものである．実際の人の動作データを用いるので，より人に近いタイミングや姿勢の動作を生成することができる．さらに，動作データの種類を増やすことによって比較的簡単に生成できる動作の種類を増やせる利点がある．しかし，生成時には使用する可能性のあるすべての動作データが参照できる状態にある必要があるので，実際に使用する動作データに対して大量の使用しない動作データを保持しておく必要がある．このため使用しないデータ分だけメモリ効率が悪くなってしまうという欠点がある．

これに対し，学習ベースの手法 [9], [10], [11], [12], [13], [14] は，事前に動作データの分布や傾向を学習することで，手などの端点の軌跡や目標物の位置のみを与えることで，それに辻褄が合う動作を予測し生成する手法である．この手法は，入力から動作を生成する生成器のみを必要とするので，検索ベースよりもメモリ効率が良い．しかし，様々な目標物や歩行の軌跡に対応するためには，それらを内包した分布や傾向を学習するように多くの動作データを事前に用意しておく必要がある．さらにこれに関連して，生成する動作データの種類を増やすためには，各種類ごとに多くの動作データを集めなければならないので，動作の種類を増やすコストが大きいことが挙げられる．

本研究で扱う協調動作は，前述したように検索ベースの手法を用いることは現実的ではない．よって，協調動作データセットを深層学習器に学習させる学習ベースの動作生成手法を選択した．

2.2 動作予測

動作予測を応用することで，動作生成を行うこともできる．動作予測は，1 人の動作データについてある部分を入力として受け取り，その先に動作がどのように続くのかを予測する技術である．これを用いて，ある動作データからその先を予測することで新しい動作データを生成すること

ができる．動作予測は，予測期間の長短とそれに伴う評価指標の違いによって 2 種類に分けることができる．

短期間の行動予測 [15], [16], [17], [18], [19] は，ごく短い範囲の未来を予測する．扱う範囲は，同じ動作がある程度続いている決定的な範囲であり，主にコンピュータビジョンにおけるトラッキングや協働ロボットが人間と衝突を避けるために用いられる．そのため評価は，予測と正解との二乗誤差を計算するのが一般的である．

長期間の行動予測 [20], [21], [22], [23], [24], [25] は，動作の不確定性を含む範囲まで予測の対象である．人間の動作には時間が進むにつれて不確定性が大きくなって行くので，予測動作と正解動作の誤差を小さくするよりむしろ状況に合致した動作の生成を目的とすることが多い．そのため評価法も，二乗誤差を用いるだけでは不適切であり，動作の質を測るような評価指標や人の手による評価が行われている．本研究においては，1 つの入力動作に対して位置とタイミングが合致した協調動作は多数考えられる．よって，本研究の評価指標は位置やタイミングの合致という数値評価と自然であるという人の手による評価の双方を行うことが望ましい．

2.3 協調動作生成の従来法

相互動作生成を行った先行研究は Vogt ら [26] が行った．この研究では，キックとパンチの動作を行う攻撃者に対して腕を用いて防御を行う防御者の協調動作を生成することを目指している．入力として，攻撃者の過去 10 フレームの姿勢を各関節点の 3 次元位置として受け取り，防御者の各関節点の 3 次元位置を予測し出力する．これを毎フレーム繰り返して，防御者の協調動作を生成している．2 層程度のパラメータ数の小さい多層パーセプトロンを用い，さらに攻撃者と防御者それぞれの姿勢の関節点の 3 次元位置に対して次元削減を行うことで，生成モデルの過学習を防いでいる．また，ヒューマノイドロボット NAO への応用も行っており，リアルタイムでユーザとの交流を実現している．

しかし，キックやパンチといった攻撃と防御の協調動作生成と比べて，本研究で生成を目指すハイタッチやフィストバンプはより位置やタイミングの制約が厳しい．なぜなら，生成するべき防御者の動作は，腕を先に構え腕全体のどこかで防御する動作であるのに対して，本研究で目指すハイタッチとフィストバンプは手のひらや拳の先を合わせる動作である．腕全体と比較して手の先といった狭い範囲を合わせる必要があることから位置の制約が厳しく，防御のように先に構えておくのではなく手を当てに行く必要があることからタイミングの制約も大きいと言える．本研究では，従来手法に対してより位置とタイミングがより厳しく求められる難しい協調動作を生成することを目標とし，フィストバンプの協調動作生成を目指した．

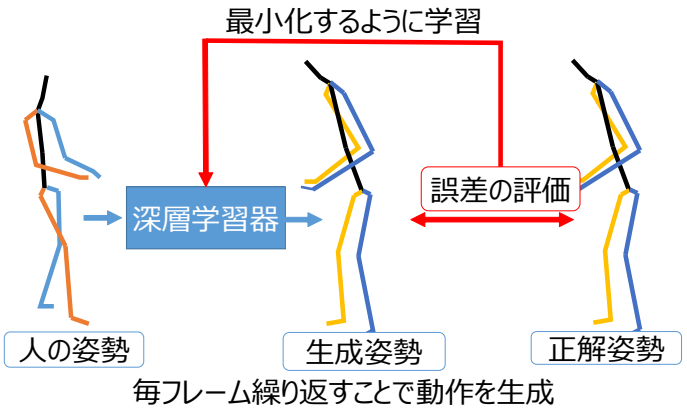

図 2 協調動作生成の流れ．動作は 3 次元姿勢の時系列集合であるので，毎フレーム，人の姿勢を深層学習器に入力し，それに対応する姿勢を生成することで動作を生成する．また，この深層学習器を用意するため，事前に 2 人のフィストバンプ動作を記録し，片方を入力動作，残りを正解動作とした．生成姿勢と正解動作の誤差が最小となるように深層学習器を学習する．

3. 協調動作生成を行う提案手法

動作は，3 次元姿勢の時系列集合である．今回は，図 2 のように毎フレーム人の姿勢を深層学習器に入力し，対応する姿勢を生成することで動作を生成する．

また，深層学習器は生成した姿勢と正解の姿勢の誤差が小さくする教師付き学習を行い用意する．学習データは，事前に 2 人のフィストバンプ動作を記録しておき，片方を入力動作，残りを正解動作とすることで作成する．以降では，提案手法である協調動作生成を行う深層学習モデルと，端点座標の正確性と関節間距離の安定性を両立する誤差評価法を述べる．

3.1 協調動作生成を行う深層学習モデル

動作などの時系列データを扱う深層学習器には，連続的な情報を考慮できる Recurrent Neural Networks（RNN）[27] が用いられる．今回は，その RNN の中で最も良い性能を示した Long Short-Term Memory（LSTM）[28] を使用した．しかし，単純に毎フレームで入力姿勢のみを LSTM に入力し，対応する姿勢を予測し生成するだけでは，各フレームで生成した姿勢の間の速度の大きさを学習することができず，フレーム間に不連続な箇所を持つ動作が生成されていた．これに対処するために，本研究では，協調動作生成のために図 3 のような深層学習モデルを提案する．

残差パスにより LSTM に各フレーム間の差分を出力するように学習することにより，各フレーム間の差分のスケールを学習することができ，不連続性が抑制される．まず毎フレームにおいて，人の 3 次元姿勢と前フレームで生成した姿勢を LSTM に入力する．そして，LSTM の出力と前フレームで生成した姿勢の和を今フレームの生成姿勢とする．生成姿勢と正解姿勢の誤差を最小化するように学習す

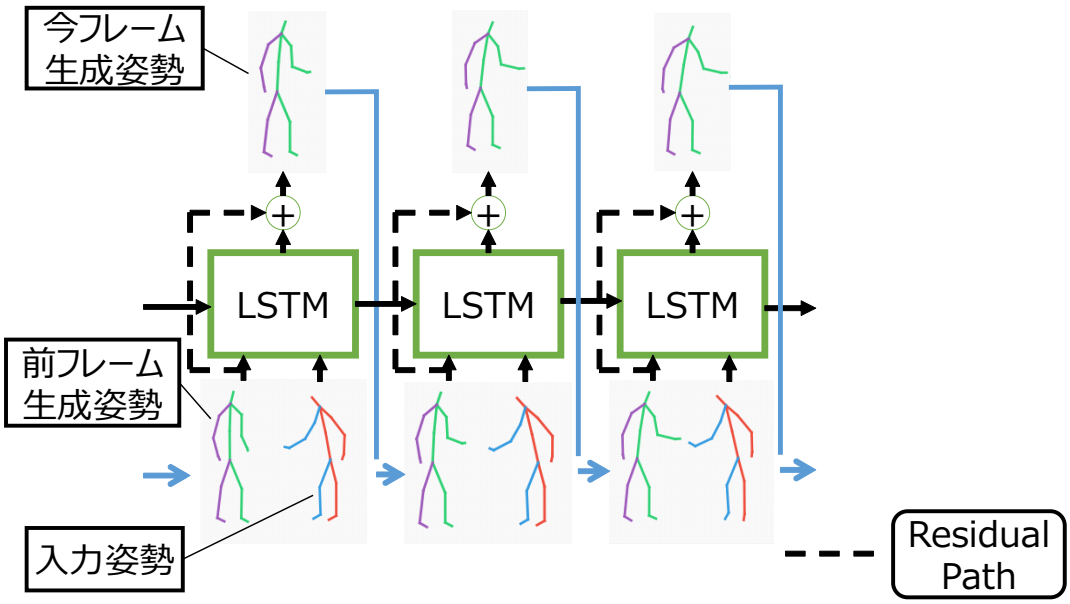

図 3 提案する深層学習モデル．毎フレーム，前フレームで生成した姿勢と入力姿勢を LSTM に入力し，今フレーム生成姿勢を得ている．生成動作の不連続性に対処するために，LSTM がフレーム間の姿勢の差分の大きさを学習するように残差パスを追加した．この残差パスによって今フレームの生成姿勢は，前フレーム生成姿勢と LSTM の出力の和となる．今フレーム生成姿勢が学習によって正解姿勢に近づくと，LSTM の出力は前フレームの生成姿勢と正解姿勢の差分を出力するようになる．これにより，LSTM がフレーム間の差分の大きさを学習でき，不連続性が抑制される．

るため，理想的には LSTM が前フレーム生成姿勢と正解姿勢の差分を出力するようになる．これにより，LSTM が各フレームの姿勢間の差分のスケールを学習でき，不連続な差分が大きい動作を生成することを抑制できる．

3.2 端点座標の正確性と関節間距離の安定性を両立する誤差評価法

本研究では，新たな誤差評価法を提案する．それを説明するために，2 つの姿勢の表現法とその学習特性について述べる．これらをまとめたものを，図 4 に示す．

まず 1 つ目の表現は，隣り合う関節間のベクトルを保持する相対ベクトル表現である．この表現では，深層学習器が各関節間ベクトルの大きさ，つまり関節間距離を学習することができ，関節間距離が安定する．しかし，各ベクトルの足し合わせにより端点座標を表現するので，誤差の累積してしまい端点座標の正確性に欠けることが分かった．これに対して，各関節の座標を保持する絶対座標表現がある．これは，各関節の座標を独立に表現するので，関節間距離が不安定となる．また，座標値を直接持つことで誤差の累積を回避できるが，座標値の絶対値が大きくなることで分散が大きくなり，それに伴い端点座標の誤差が大きくなることが分かった．よって，この 2 つの姿勢表現それぞれで誤差を最小化する学習を行うことでは，端点座標の正確性と関節間距離の安定性とを両立することはできない．

これに対処するために，図 5 のような誤差評価法を提案する．まず，深層学習器に相対ベクトル表現の姿勢を生成させる．この姿勢を足し合わせにより絶対座標表現へ変換する．その後，相対ベクトル表現と絶対座標表現両方での誤差を評価する．深層学習器に，相対ベクトル表現の姿勢を学習させることで，絶対座標表現で発生した各ベクトルの分散の大きさによる誤差を抑制し，相対ベクトルを学習することで関節間距離の安定性を得ることができる．また，絶対座標まで足し合わせたもので誤差を評価しそれを最小化するので，足し合わせ後の誤差を最小化でき，誤差の累積を予測できる．これにより，端点座標の正確性と関節間距離の安定性を両立した学習ができる．

4. 実験

今回は，2 人のフィストバンプ動作データセットを新たに作成し，提案手法の有用性を確かめた．フィストバンプ動作は kinect v2 センサー [29]（以下 kinect）を用いて撮影した．これを使用した理由は，kinect がモーションキャプチャシステムと比べて，専用スーツを着なくても動作を撮影できるという利便性から，より多くの人を対象にして撮影できることでデータセット内の動作の多様性を上げること，専用スーツを着ないことから動きに制限がなく日常の動作そのままを記録できるからである．kinect を用いた動作撮影の俯瞰図を，図 6 に示す．また kinect は，床から 100cm の台に設置した．データセット内の動作の多様性を確保し対応できる動作の範囲を広げるために，動作を行う

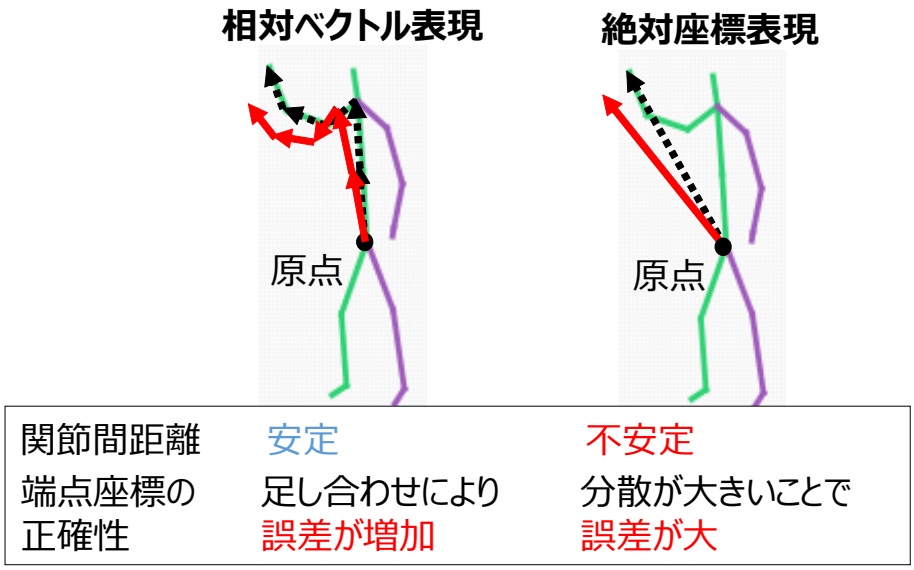

図 4 2 通りの姿勢表現とその学習特性．各姿勢表現は破線で表されるベクトルを持ち，図中のような学習特性を持つ．

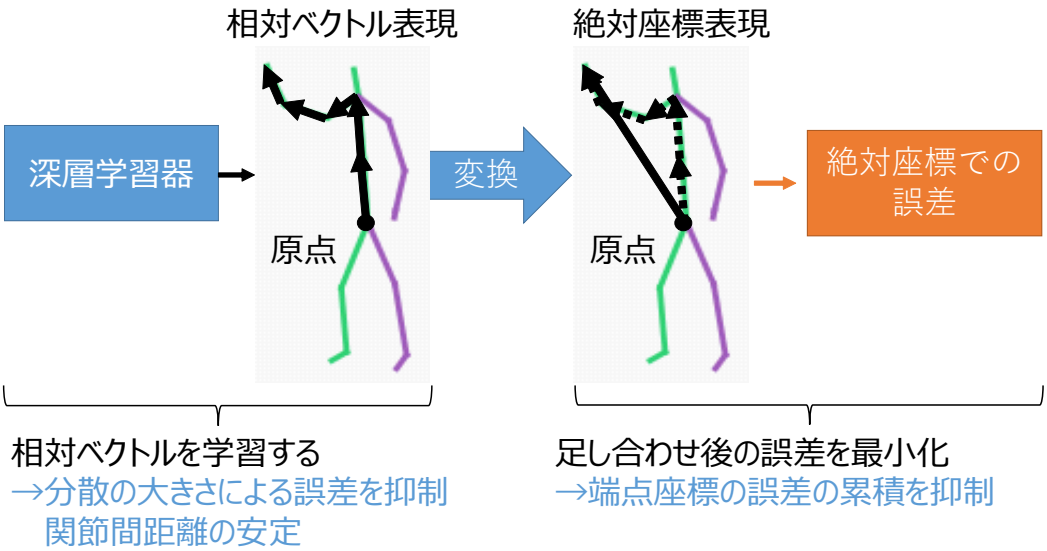

図 5 端点座標の正確性と関節間距離の安定性を両立する誤差評価法．

人の多様性，フィストバンプに使用する手の多様性，高さの多様性をもたせた．動作する人の多様性を持たせるために，22 人の組み合わせで動作を撮影した．使用する手の多様性をもたせるために，以下の 3 つを変えて動作の撮影を行った．

(1) 2 人の動作者の組み合わせ
(2) 2 人の使用する手の組み合わせ
(3) 手を合わせる高さ

1 つ目では，22 人の動作者を組み合わせて撮影を行った．2 つ目では，2 人の左右の手の組み合わせ 4 つにを両手を含めて 5 種類の動作を記録した．3 つ目では，手の組み合わせに加えて，手を合わせる高さを上中下段の 3 種類に変えて撮影を行った．これにより，636 個のフィストバンプ動作を得た．得られたデータセットを，3 : 1 の比率で 477 個の学習データと 159 個のテストデータに分割した．そして，各手法において学習データを用いて学習を行い，テストデータで生成した動作を評価した．

生成した動作を評価するために，以下の 4 つの評価指標を用いた．

(1) 二乗平均平方根誤差

$$RMSE = \sqrt{\frac{1}{FJ}\sum_{f}^{F}\sum_{j}^{J} \| \hat{\mathbf{y}}_{f,j} - \mathbf{y}_{f,j} \|^2}$$

(2) 関節間距離の標準偏差

表 1 各手法による生成動作を 4 つの評価基準によって評価した．その結果すべての基準において提案手法が最も良い結果を残している．

	RMSE[cm]	関節間距離の標準偏差 [cm]	接触フレーム誤差	接触時の端点座標誤差 [cm]
従来法	8.18	4.91	12.74	54.4
深層学習モデル＋絶対座標表現	3.96	3.58	4.45	17.6
深層学習モデル＋相対ベクトル表現	3.92	3.32	4.39	16.3
提案手法	**3.64**	**3.10**	**3.69**	**13.4**

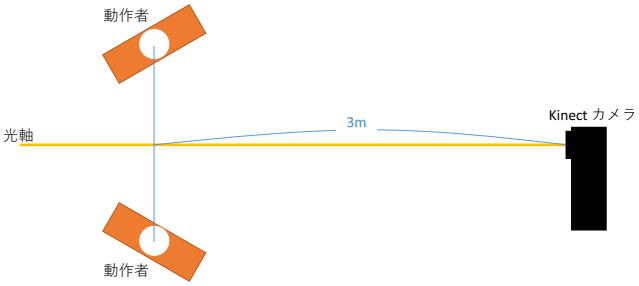

図 6 動作データ作成時のカメラ及び動作者の俯瞰図．kinect v2 の人物検出範囲は 0.5〜4.5m である．その中で，2 人の動作が収まるように kinect カメラを動作者から 3m 離して動作を撮影した．

$$\text{関節間距離の標準偏差} = \sqrt{\frac{1}{FL}\sum_{f=1}^{F}\sum_{j=1}^{J} -1(d_{f,j} - \bar{d}_j)^2}$$

$$\text{ただし}, d_{f,j} = \| \mathbf{y}_{f,j+1} - \mathbf{y}_{f,j} \|$$

$$\bar{d}_j = \frac{1}{F}\sum_{f=1}^{F} d_{f,j}$$

(3) 接触フレーム誤差

$$\text{接触フレーム誤差} = | \hat{f}_{col} - f_{col} |$$

$$\text{ただし}, f_{col} = \arg\min_{f} \| \mathbf{x}_{f,hand} - \mathbf{y}_{f,hand} \|$$

$$\hat{f}_{col} = \arg\min_{f} \| \mathbf{x}_{f,hand} - \hat{\mathbf{y}}_{f,hand} \|$$

(4) 接触時の端点座標誤差

$$\text{接触時の端点座標誤差} = \| \hat{\mathbf{y}}_{f_{col},hand} - \mathbf{y}_{f_{col},hand} \|$$

1 つ目は，正解動作の再現度を表す二乗平均平方根誤差（RMSE）である．2 つ目は，関節間距離の安定性を示す関節間距離の標準偏差である．3 つ目と 4 つ目は，端点座標の正確性を評価している，手の接触フレーム誤差と接触時の端点座標誤差である．接触フレーム誤差は，入力動作の手の座標と生成動作，正解動作の手の座標の距離が最も小さくなったフレームを接触フレームとする．この生成動作の接触フレームと正解動作の接触フレームの差が接触フレーム誤差である．これにより，生成動作のタイミングを評価できる．また，接触時の端点座標誤差は，正解動作の接触フレームにおける正解動作の手の端点座標と生成動作の手の端点座標の距離である．これにより，接触時の端点座標の正確性を評価できる．これら 4 つの評価指標はすべて小さいほど良い値である．

実験結果を表 1 に示す．また，正解動作と各手法によって生成された動作の接触フレーム周辺の姿勢を図 7 に示す．従来法は，すべてのフレームで同じ姿勢を出力していた．これは，従来法が多層パーセプトロンを用いて各フレームの姿勢を独立に出力していることで，連続的な特徴を捉えることができなくなり，その結果平均的な姿勢を出力してしまったためであると考える．絶対座標表現での学習では，左手（紫）の関節間距離の長さが伸び縮みしていることが分かる．これでは，不自然な姿勢である．相対ベクトル表現での学習では，右手（黄緑）の端点座標の誤差が絶対座標表現と比べて大きい．これでは，入力動作との時空間同期にに欠けてしまう．これと比べて，提案手法では大きな関節間距離の変化は見られず，さらに端点座標の誤差も他の手法と比べて小さい．

5. まとめと今後の課題

本研究では，仮想空間内のキャラクターやヒューマノイドロボットが人と高度な交流を行うために，人の動作からそれに対応する動作を生成する手法を提案した．その結果，従来法と比べて二乗平均平方根誤差を 8.18cm から 3.64cm へ低減できた．また，関節間距離の標準偏差，接触フレーム誤差，接触フレームでの端点座標誤差についても従来法よりも良い結果を出力している．これらより，端点座標が正確で関節間距離が安定した協調動作を生成することができた．

今後の課題は 2 つ挙げられる．1 つ目は，深層学習器を学習させるために生成動作と正解動作の誤差を最小化するようにのみ学習していることである．人間がフィストバンプ動作に対応する際は，左右の手どちらで相手の手に対応

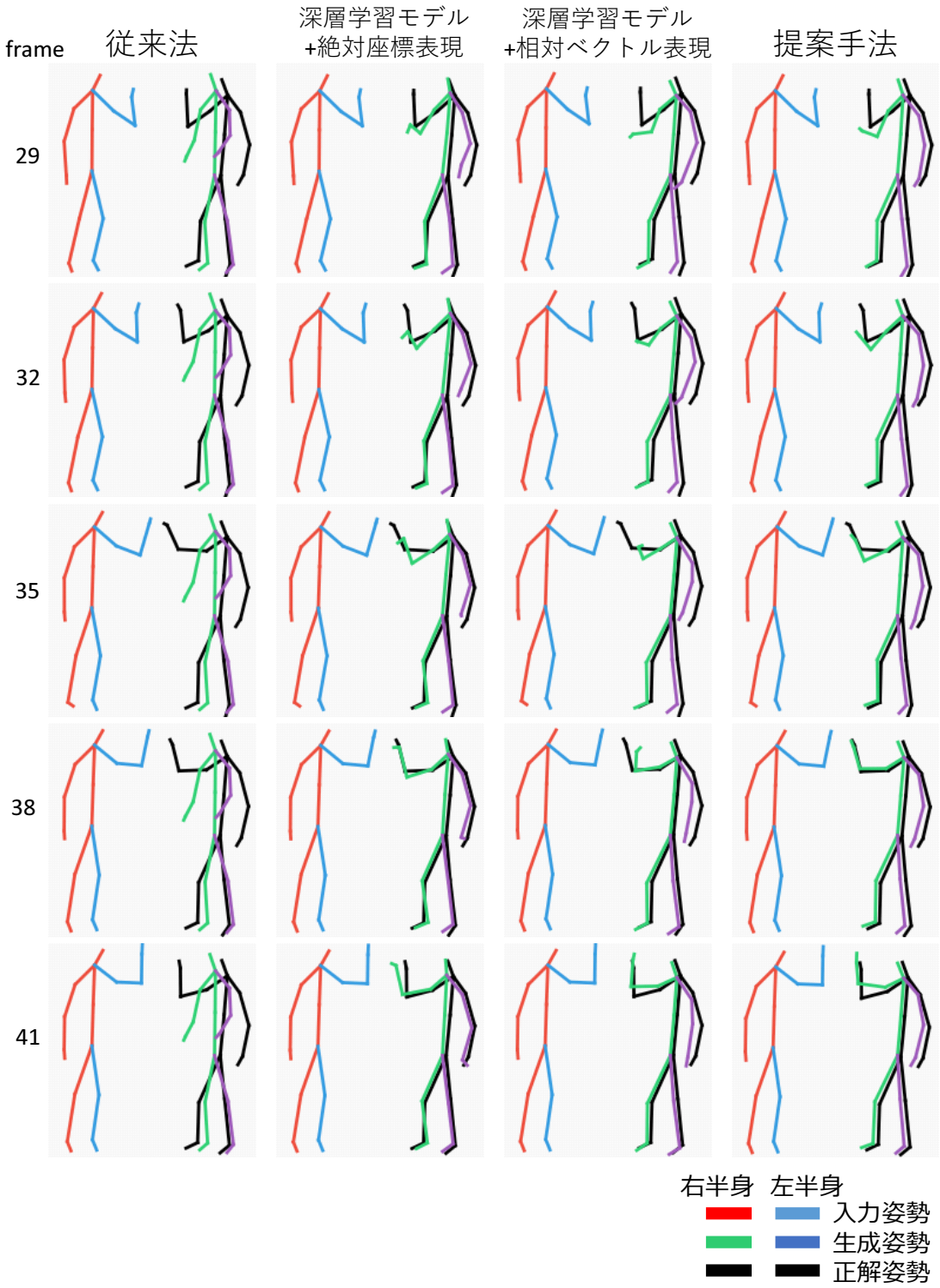

図 7 各手法によって生成された協調動作．各列が各手法で生成した動作と対応している．赤と青の表現が入力姿勢，黄緑と紫の表現が生成姿勢，黒の表現が正解姿勢である．従来法は静止した動作を生成している．絶対座標表現での学習は左手（紫）の手首から手の先の距離がが 29 フレーム目から 41 フレーム目まで不自然な伸び縮みをしており，関節感距離が不安定な動作を生成している．相対ベクトル表現での学習は，誤差の累積によって右手（緑）の座標の誤差が大きくなってしまい，接触フレーム（35 フレーム）で右手が接触できておらず，時空間的な同期が取れていない動作を生成している．これらに対して提案手法は，関節感距離の安定と誤差の累積の抑制に成功し，他手法と比べて最も正解動作に近く最も自然な協調動作を生成している．

ⓒ 2019 Information Processing Society of Japan

するかなどの自由度がある．この自由度は，誤差の最小化では表現しきれず，動作のバリエーションを制限してしまう．よって，動作認識の結果を損失関数に加えるなどの動作の自由度を持たせられるような誤差評価が求められる．2 つ目は，現在の手法では協調動作 1 種類を生成するために大量のデータが必要となることである．今回はフィストバンプ動作を生成するために 22 人から 636 個の動作を撮影した．人が行う動作の種類は膨大であり，1 種類の動作を生成するために多くの人から動作を記録することは効率的ではない．よって，動作というドメインに最適化されたデータ拡張を行うことによって，より少ない協調動作の事例からの動作生成を行うことを考えている．

参考文献

[1] SoftBankRobotics: Nao®.
[2] SoftBankRobotics: Pepper®.
[3] Tolani, D., Goswami, A. and Badler, N. I.: Real-time inverse kinematics techniques for anthropomorphic limbs, *Graphical models*, Vol. 62, No. 5, pp. 353–388 (2000).
[4] Grochow, K., Martin, S. L., Hertzmann, A. and Popović, Z.: Style-based inverse kinematics, *ACM transactions on graphics* , Vol. 23, No. 3, pp. 522–531 (2004).
[5] Aristidou, A. and Lasenby, J.: FABRIK: A fast, iterative solver for the Inverse Kinematics problem, *Graphical Models*, Vol. 73, No. 5, pp. 243–260 (2011).
[6] Yang, Y., Ivan, V. and Vijayakumar, S.: Real-time motion adaptation using relative distance space representation, *2015 International Conference on Advanced Robotics (ICAR)*, Vol. 1, pp. 21–27 (online), (2015).
[7] Holden, D., Saito, J. and Komura, T.: A deep learning framework for character motion synthesis and editing, *ACM Transactions on Graphics*, Vol. 35, No. 4, pp. 1–11 (online), (2016).
[8] Ye, G. and Alterovitz, R.: guided motion planning, In *Robotics research*, pp. 291–307 (2011).
[9] Peng, X. B., Berseth, G., Yin, K. and Van De Panne, M.: DeepLoco: Dynamic Locomotion Skills Using Hierarchical Deep Reinforcement Learning, *ACM Transactions on Graphics*, Vol. 36, No. 4, pp. 1–13 (online), (2017).
[10] Holden, D., Komura, T. and Saito, J.: Phase-functioned neural networks for character control, *ACM Transactions on Graphics*, Vol. 36, No. 4, pp. 1–13 (online), (2017).
[11] Kwon, T. and Hodgins, J. K.: Momentum-mapped Inverted Pendulum Models for Controlling Dynamic Human Motions A Control System for Dynamic Human Motions using a Momentum-mapped Inverted Pendulum Model Captured from Reference Motion Capture Sequences, *ACM Transactions on Graphics*, Vol. 28, No. 10 (2017).
[12] Huang, H., He, R., Sun, Z., Tan, T. et al.: Introvae: Introspective variational autoencoders for photographic image synthesis, *Advances in Neural Information Processing Systems*, pp. 52–63 (2018).
[13] Peng, X. B., Abbeel, P., Levine, S. and van de Panne, M.: DeepMimic: Example-Guided Deep Reinforcement Learning of Physics-Based Character Skills, *ACM Transactions on Graphics*, Vol. 37, No. 4 (2018).
[14] Liu, L. and Hodgins, J.: Learning Basketball Dribbling Skills Using Trajectory Optimization and Deep Reinforcement Learning, *ACM Transactions on Graphics*, Vol. 37, No. 4, p. 14 (online), (2018).
[15] Bütepage, J., Kjellström, H. and Kragic, D.: Anticipating many futures: Online human motion prediction and synthesis for human-robot collaboration, *arXiv preprint arXiv:1702.08212* (2017).
[16] Perez-D'Arpino, C. and Shah, J. A.: Fast target prediction of human reaching motion for cooperative human-robot manipulation tasks using time series classification, *Proceedings - IEEE International Conference on Robotics and Automation*, Vol. 2015-June, No. June, pp. 6175–6182 (online), (2015).
[17] Pérez-D'Arpino, C. and Shah, J. A.: Fast motion prediction for collaborative robotics, *IJCAI International Joint Conference on Artificial Intelligence*, Vol. 2016-Janua, pp. 3988–3989 (2016).
[18] Schydlo, P., Rakovic, M., Jamone, L. and Santos-Victor, J.: Anticipation in Human-Robot Cooperation: A recurrent neural network approach for multiple action sequences prediction, *IEEE International Conference on Robotics and Automation*, pp. 1–6 (2018).
[19] Gui, L. Y. and Wang, Y. X. and Liang, X. and Moura, J. M.: Adversarial geometry-aware human motion prediction, *European Conference on Computer Vision*, pp. 786–803 (2018).
[20] Fragkiadaki, K., Levine, S., Felsen, P. and Malik, J.: Recurrent Network Models for Human Dynamics, *2015 IEEE International Conference on Computer Vision*, pp. 4346–4354 (online), (2015).
[21] Barsoum, E., Kender, J. and Liu, Z.: HP-GAN: Probabilistic 3D human motion prediction via GAN, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1418–1427 (2018).
[22] Martinez, J., Black, M. J. and Romero, J.: On human motion prediction using recurrent neural networks, *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp. 4674–4683 (2017).
[23] Li, Z., Zhou, Y., Xiao, S., He, C., Huang, Z. and Li, H.: Auto-conditioned recurrent networks for extended complex human motion synthesis, *arXiv preprint arXiv:1707.05363* (2017).
[24] Tang, Y., Ma, L., Liu, W. and Zheng, W.: Long-Term Human Motion Prediction by Modeling Motion Context and Enhancing Motion Dynamic, *arXiv preprint arXiv:1805.02513* (2018).
[25] Pavllo, D., Grangier, D. and Auli, M.: QuaterNet: A Quaternion-based Recurrent Model for Human Motion, *arXiv preprint arXiv:1805.06485* (2018).
[26] Vogt, D., Amor, H. and Berger, E.: Learning Two-Person Interaction Models for Responsive Synthetic Humanoids, *Journal of Virtual Reality and Broadcasting*, Vol. 11, No. 1 (online), available from ⟨http://www.jvrb.org/past-issues/11.2014/3856/1120141.pdf⟩ (2014).
[27] Goodfellow, I., Bengio, Y. and Courville, A.: Deep Learning, *Nature*, Vol. 521, No. 7553, p. 800 (online), DOI: 10.1038/nmeth.3707 (2016).
[28] Hochreiter, S. and Schmidhuber, J.: Long short-term memory, *Neural computation*, Vol. 9, No. 8, pp. 1735–1780 (1997).
[29] Zhang, Z.: Microsoft kinect sensor and its effect, *IEEE multimedia*, Vol. 19, No. 2, pp. 4–10 (2012).

ⓒ 2019 Information Processing Society of Japan