第 27 回日本バーチャルリアリティ学会大会論文集（2022 年 9 月）

ＮＮ内挿補間を用いた低サンプリングレート関節点の高サンプリングレート化

Generation of High-Sampling-Rate Motion Capture Data
from Low-Sampling-Rate Data based on NN Interpolation

飯田航平 [1)]，清水創太 [1)]，奥野哲史 [1)]
Kohei IIDA, Sota SHIMIZU, and Satoshi OKUNO

1) 芝浦工業大学大学院電気電子情報工学専攻 （〒108-8548 東京都港区芝浦 3-9-14, ma22012@shibaura-it.ac.jp）

概要： 本発表では，メタバース等の時代のトレンドワードに代表される仮想現実空間において，対人インタフェースとしての役割を担う 3D アバターを違和感なく滑らかに動かすことを目標としている．比較的安価にモーションキャプチャーを実現できるものの，計算コストの大きさから低サンプリングレートである OpenPose のような定点シングルカメラベースの関節点の時系列データを NN 内挿補間によってリアルタイムで高サンプリングレート化する手法について述べる．

キーワード：低サンプリングレートモーションキャプチャーデータ，高サンプリングレート化，ニューラルネットワーク内挿補間，　3D アバター

1. はじめに

メタバースは時代のトレンドワードであり，おそらく若い世代を中心に一度はニュース等で耳にしたことがあるのではないであろうか？この仮想現実空間技術のキーワードの一般社会への浸透は，近年の VR の需要が以前にも増して高まっていることの証明であるといえよう．ゲームやエンターテイメント業界に限らず，その広がりは医療や不動産，ショッピング，観光，スポーツ，教育にまで裾野を広げている[1]．VR 市場の動向を示す業界マップの 2019 年版を図 1 に示す．VR 市場に参画している企業の多種・多様さが本図より見て取れる．また，株式会社リプロネクストの 20～40 代に対しての VR に関するアンケート調査が興味深い結果を示している[2]．当該の会社は実写 VR コンテンツや，VR ゴーグル制作，インターネット広告事業等を手掛ける企業である．その調査結果では，全体の約 8 割が，コロナ禍で VR を耳にする機会が増えたと回答している．VR 技術の進歩には目覚ましいものがあるが，著者らは，モーションキャプチャー技術が VR の市場規模をさらに拡大させる重要な要素技術であると捉えている．実際に，エンターテイメント分野への応用以外では，ヘルスケアの分野の AR 手術ガイダンスシステムのトレーニング等への導入の報告がある[3]．このように，多種多様な分野でモーションキャプチャー技術が活用されるようになっている．しかし，現在の技術では精度が良く違和感のない人間の動作のトラッキングを行うには高価な機器や設備を必要とする．この側面は企業等の大型スタジオと異なり，数多くの人員の確保や前述の高価な機材を用意するのが難しい個人ユーザーにとっては頭の痛い問題となる．

さて，手軽にモーションキャプチャーをする方法として OpenPose のような定点カメラからの画像を用いる方法が知られている[4]-[6]．これらの手法では，映像から関節点を定めてスケルトンをフィットさせることで，簡便に各関節点推定を行い，その時系列データを取得している．ここでは，単眼の可視光カメラ 1 台さえあれば良く，近赤外マーカーや複数台の近赤外カメラ，決して狭くない撮影空間を必要としないことがメリットである．しかし，その一方で，スケルトン当てはめの演算コストが非常に大きく，高性能化した CPU や GPU による並列演算技術により処理の高速化が著しい現在のコンピューター技術をもってしても高サンプリングレートのモーションキャプチャーデータを取得するのは困難である．その結果，得られたデータに 3D アバターを連動させたとき，その動きは実にカクカクしたものとなる．こうした背景から，線形補間などの計算コストの低い補間手法を用いることで，低サンプリングレートで得られた関節点データから，リアルタイムで低コストに高サンプリングレート化するという着想に到った[7]．また，この技術は位置精度の高くない時系列関節点データの補正やある時刻において欠損のあるデータの補間にも役立つと考えた．本研究の最終的な目標は，カメラを用いて手軽に取得した低サンプリングレートの関節点データから，リアルタイムで違和感なく滑らかな高サンプリングレートデータを生成することである．

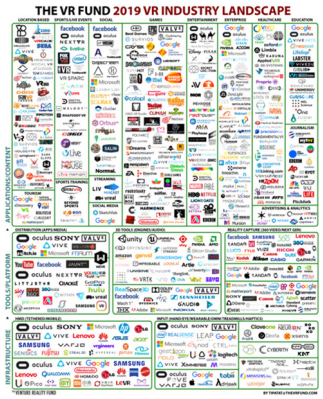

図 1　VR 市場の動向を示す業界マップの 2019 年版[1]

2. 関連・先行研究

単眼カメラを用いて姿勢推定をする研究は多くされている[4]-[6]．その中でも，OpenPose はカーネギーメロン大学(CMU)の Zhe Cao らが発表した 1 台の定点可視光カメラを用いた簡便なモーションキャプチャー手法である．本手法では，得られた動画から切り出された各静止画に対して keypoint(特徴点)の検出と keypoint どうしの関係の推定に深層学習を用いてスケルトンを当てはめて姿勢の推定を行い，連続する姿勢推定をリアルタイムで行うことができる[4]．しかし，かなり計算コストが高いため，高性能なコンピュータが必要となり，リアルタイム処理の場合 3Hz から 10Hz にまでフレームレートが下がってしまい，上述のように 3D アバターに応用したとき，カクカクとしたぎこちない動きとなってしまう．OpenPose を動画に適用して，関節点の検出が失敗している箇所を検出失敗フレームとみなし，線形補間を用いて座標値を補間及び補正を行うことで骨格推定精度を改善する研究についての報告がある[8]．補間技術については，動画像についてのフレーム補間が多く研究されている[9]-[11]．これらの手法をフレーム補間以外に用いた研究に関して，深層学習を用いた時系列補間技術の非画像データへの応用時の適用性評価の研究についての報告がある[11]．この研究では深層学習を用いた動画の内挿補間技術を元に，様々な非画像データを対象とした時系列データの内挿補間への応用について検討している．一例として気象レーダーにおいて観測された雨量データに対する時系列補間の適用結果について，深層学習に基づく推論値と位置補間，速度補間の手法を比較して紹介している．本研究では，雨量データではなく低サンプリングレートのモーションデータの高サンプリングレート化のために，機械学習と線形補間，速度補間を行う．

3. 高サンプリング化手法

低サンプリングレートの関節点データから，精度の良い，滑らかな高サンプリングレートのモーションデータを生成するために，本研究では以下の 3 つの手法のメリット・デメリットを比較する．すなわち，計算コストの低い①線形補間，②速度補間とともに，計算コストはやや高いが柔軟な動きに対応できる③ニューラルネットワーク補間(以下 NN 補間)を用い，リアルタイムに滑らかなアバター表示を行うための適用方法を考察する．以下にこれらの 3 種類の内挿補間手法による低サンプリングレートモーションキャプチャーデータの高サンプリングレート化手法について説明する．

3.1 線形補間と速度補間

これら 2 つの補間手法に関しては，よく知られる一般的な手法を用いた[11]．線形補間は連続する 2 フレームの関節点座標値を用いて中間フレームの内挿補間値を算出する．速度補間は連続する 3 フレームの関節点座標値から連続する 2 フレームの関節点の速度を算出し，対応する座標値とともに中間フレームの内挿補間値を算出する．

3.2 NN 補間

本稿では，入力層と出力層に中間層が 2 層の 3 層のニューラルネットワークを用いて学習を行い中間フレームの内挿補間値を予測する．ここでは，連続する前後 2 フレームの座標値である P_t と P_{t+dt} の集合を入力とする．このとき，P_t とは各関節点の頭から右手までの 10 関節点の座標値$(x_i, y_i, z_i)(i = 0, 1, \ldots, 9)$である．$P_t$ と P_{t+dt} の中間フレームの 10 関節点の座標値を出力とする．各中間層のノード数は 100 とした．

4. 比較検証実験

4.1 検証方法

本稿では，低サンプリングレートデータから 80Hz の高サンプリングレート化の比較・検証のための実験を行う．データの収集は図 2 に示す VRHMD である HTC VIVE Pro Eye と被験者の関節点を計測するための VIVE Tracker を用

図 2　検証実験に用いたモーションキャプチャー装置

表 1　データセット

動作の詳細	時間	データ数
ゆっくり歩く	46秒	3689
ジョギング，素早く手を振る	58秒	4639
ラジオ体操1	182秒	14602
ラジオ体操2	191秒	15322
走る，ドリル(反復横跳び)	54秒	4384

いて行う．①まず，10 関節点のモーションキャプチャーデータの計測を行った．このとき，一旦，時間間隔が均等な 80Hz のモーションキャプチャーデータを計測し，間引くことで 10Hz や 5Hz などの低サンプリングレートデータを生成した．データセットは，様々なモーションデータを取得するために表 1 に示す，「歩く」，「ジョギング」，「跳ぶ」，「体を傾ける」，「回す」，「曲げる」，「回転する」，「しゃがむ」という動作が含まれる動作データを計測した．②次に，線形補間，速度補間，NN 補間の 3 つの内挿補間手法を用いて，低サンプリングレートデータから，80Hz の高サンプリングレートデータを生成する．③高サンプリングされたデータから間引かれる前のオリジナルの 80Hz のデータを真値として最小二乗誤差を算出し精度評価を行う．

4.2 実験結果と考察

図 3 は，10Hz と 5Hz それぞれのゆっくり歩く動作が含まれている学習用データに対して 2 倍の高サンプリングレート化を行ったときの平均二乗誤差の値であり，横軸をフレーム数としてプロットしている．赤線が線形補間，緑線が速度補間，青線が NN 補間の結果となっている．この結果から，速度補間は誤差の振れ幅が大きく結果が極端に悪いことがわかる．この結果から以降は速度補間を省いた結果を示す．図 4 は，10Hz と 5Hz それぞれの素早く走る動作が含まれている学習用データに対して 2 倍の高サンプリングレート化を行ったときの平均二乗誤差の値である．図 5 は，10Hz と 5Hz それぞれの素早く走る動作が含まれている評価用データに対して 80Hz への高サンプリングレート化を行ったときの平均二乗誤差の値である．この結果を一見すると，NN 補間にはアドバンテージが見受けられないように感じられる．しかし，これは 10 関節点すべての平均二乗誤差を算出した結果である．読者らも直観的に理解できるように，人間の動作においては体幹部のように素早い動作の中でも比較的ゆっくり動く部位もあれば，手足等の先端部のようにより速く動作する部位も存在する．図 6 が示すように例えば，左足先端部の推定値はより速度の速い区間において NN 補間が線形補間の精度を上回っている箇所があることがわかる．

これらから，線形補間は動きが小さいときや時間間隔が短い場合に概ね全体的に推定精度が良いが，より素早い動きや時間間隔が長い場合に関しては NN 補間の方が精度が良くなることがあることがわかった．これは，後者の条件での内挿補間値の推定において，直線的な補間である線形補間よりも，円弧補間をカバーできる NN 補間が有効であることを示している．また，速度補間は誤差が大きく，本実験ではメリットを見出すことが出来なかった．このように，本稿における実験条件では，全体として線形補間の推定結果が予想以上に優れており，NN 補間に関してはある一定の速度を超えたときに線形補間を上回ることがあるという結果に留まった．このことは，線形補間をベースに，線形補間が苦手とする動作に対して NN 補間をハイブリッドして使用することの有効性を示している．同時に円

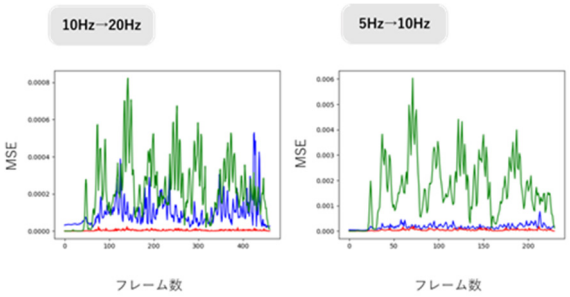

図 3　ゆっくり動く動作の学習用データからの予測結果

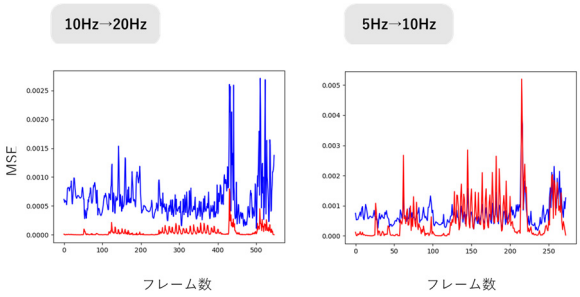

図 4　素早く走る動作を含む学習用データからの予測結果

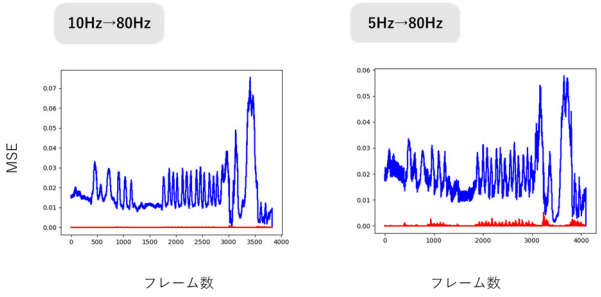

図 5　素早く走る動作を含む評価用データからの予測結果

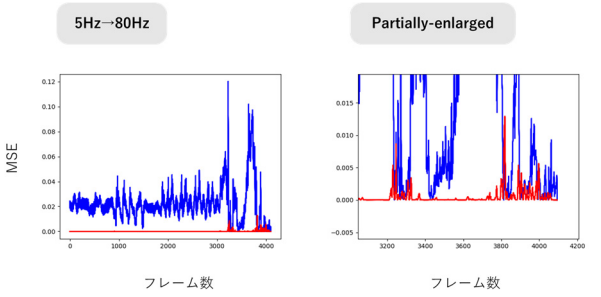

図 6　素早く走る動作時の左足先端部の評価用データからの予測結果

弧補間を要する内挿補間が求められるより低いサンプリングレートの条件下において，NN 補間のみによる推定手法が必要となる可能性を強く示唆している．

5. おわりに

本研究では，80Hz で取得した 10 関節点モーションキャプチャーデータを間引いた低サンプリングレートデータに対して，NN 補間と線形補間，速度補間を用いて，高サンプリングレート化を行い，比較・検証及び評価・考察を行った．結果として，線形補間をベースに，線形補間が苦手とする動作に対して NN 補間をハイブリッドして使用することの有効性が示めされた．また，同時に NN 補間のみ

による推定手法が必要となる条件が存在することが示された．本稿における NN 補間は 3 層の全結合型 NN による回帰分析器によって実装されたが，円弧補間を行う際には，このモデルであっても線形補間を超えるパフォーマンスを得ることが出来た．今後は，この結果を元に NN 補間回帰分析器のモデルの改良を行い，以下のことに着眼して研究を進めて行きたい：①必然的にサンプリングレートを低くする欠損データが存在する場合の推定精度の向上，②人間の身体構造の拘束条件を加味した CNN 及び LSTM のようなより強力な深層学習モデルの導入，③リアルタイム処理を可能とするための計算コストの改善，④未来の動作を予測できる外挿補完回帰分析器への発展．

謝辞

本研究の一部は科研費 No.18K04055 と No.21K03983 に基づいて実施されている．本研究の遂行に当たり，技術面のみならず多くの点でご助言，ご協力頂いた芝浦工業大学デザイン工学科人間支援知能ロボティクス研究室の皆さまに心より御礼申し上げます．

参考文献

[1] MoguraVR News(online), VR 業界マップ 2019 年版が公開成長が顕著な分野は?, https://www.moguravr.com/vr-industry-map-2019, 2019/5/23 (2021/06/30 閲覧).

[2] PRTIMES, VR についての意識調査を実施!, 株式会社リプロネクスト, https://prtimes.jp/main/html/rd/p/000000010.000064851.html, 2021/4/28 (2021/12/27 閲覧)

[3] 杉本，谷口，新城，XR(VR・AR・MR)によるテレイグジスタンス・超臨場感コミュニケーションと遠隔医療・手術シミュレーション・トレーニング，バイオメカニズム学会誌，Vol.43, No. 1, pp. 35-40 (2019)

[4] Z. Cao, T. Simon, S. Wei, Y. Sheikh, Realtime multi-person 2D pose estimation using part affinity fields, Proc. of CVPR (2017)

[5] A. Toshev and C. Szegedy, Deeppose: Human pose estimation via deep neural networks, Proc. of CVPR (2014).

[6] X. Peng, Z. Tang, et.al., Jointly optimize data augmentation and network training: Adversarial data augmentation in human pose estimation, Proc. of CVPR (2018).

[7] 飯田，清水，奥野，機械学習を用いた低サンプリングレート関節点の高サンプリングレート化，電気学会産業計測制御研究会技術資料, IIC-21-053 (2021)

[8] 山川，石川，渡辺，時系列相関性を用いた姿勢推定モデルの精度向上,情報処理学会第 82 回全国大会講演論文集, pp. 249-250 (2020).

[9] G. Long, L. Kneip, et.al., Learning image matching by simply watching video. Proc. of ECCV (2016)

[10] J. Van Amersfoort, W. Shi, et.al., Frame interpolation with multi-scale deep loss functions and generative adversarial networks, Proc. of CVPR (2017)

[11] S. Niklaus，L. Mai，F. Liu，Video Frame Interpolation via Adaptive Separable Convolution，Proc. of ICCV，pp. 261-270 (2017)

[12] 川嶋，小堀，深層学習を用いた時系列補間技術の非画像データへの適用性評価, MSS 技法, Vol.29 (2019).

[13] Y. Yuan, K. Kitani, Ego-Pose Estimation and Forecasting as Real-Time PD Control, Proc. of ICCV (2019)