

ユーザ姿勢と VR 環境情報を考慮した CG アバタの操作支援

山田 晃嗣^{† a)} 仲 泰誠^{† b)} 外山 史^{† c)} 森 博志^{† d)}宇都宮大学[†]

1. はじめに

CG アバタは VR におけるユーザの代替映像として用いられる。中でもモーションキャプチャを利用したアバタの操作では、ユーザの動作入力からアバタに反映されることから没入感の喪失を抑えた操作が期待できる。

操作時に高い没入感を得るためにはユーザが操作の際に意図した動作と同質の動作をアバタが実行できることが重要となる。そのためユーザには構成したいアバタの動作と同様の詳細な演技が求められる。

このような問題に対し、部分的な姿勢入力からユーザの操作意図を反映するアバタの動作構成手法 [1] では、全身の姿勢情報が取得できない場合においても部分的な姿勢入力からユーザが意図する全身姿勢を推定しアバタに反映することを可能にしている。しかし、VR 環境においてアバタが置かれている状況については考慮されおらず、同様の姿勢入力においても VR 環境に依存して異なる操作意図がある場合にはアバタの動作と VR 環境との整合性を保つことが難しい。

そこで本稿ではユーザの姿勢入力と CG アバタが置かれている VR 環境における周囲の環境情報を基に、機械学習を用いてユーザの意図を反映し且つ VR 環境と整合性のあるアバタ動作を構成する手法を提案する。VR 環境におけるアバタの全身動作と、アバタの周囲の環境情報および動作の実行を意図したユーザの部分的な姿勢入力とを訓練データとして学習することで、ユーザの姿勢入力から VR 環境との整合性が保持されたアバタの全身動作を構成する。これによりユーザはアバタと VR 環境との整合性を意識した詳細な演技による操作を必要とすることなく操作意図を反映したアバタの全身動作を得られることが期待できる。

2. 提案手法

2.1. 概要

図 1 に提案手法の概要を示す。本手法ではユーザの任意の身体部位の動作情報とアバタの周囲の環境情報を入力として、VR 環境との整合性が

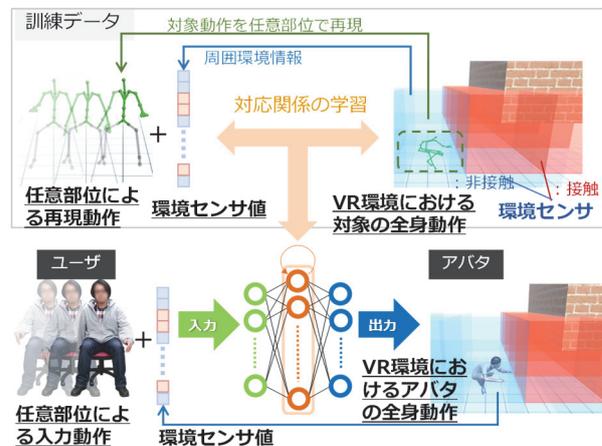


図 1 提案手法の概要

保持されたアバタの全身動作を構成する。アバタ動作構成モデルは、リカレントニューラルネットワークを用いて構築し、対象の VR 環境におけるアバタの全身の動作データを教師データ、その動作の操作を意図したユーザの任意の身体部位による姿勢情報とアバタの環境センサ情報を入力データとした教師あり学習により、任意部位および周囲の環境情報の時間的特徴と全身の動作との対応関係を学習する。

本稿ではユーザが操作を行う際の任意の身体部位として、座位姿勢でアバタを操作することを想定し上半身を対象とした。またアバタが置かれている VR 環境の周囲の情報の取得には [2] を参考に、アバタモデルの周囲に配置された立方体領域の集合である環境センサを定義し、各立方体領域と環境物の接触判定結果を環境センサ情報とした。

2.2. 訓練データの作成

学習に用いる訓練データセットは次の手順で作成する。

- (1) 対象となる VR 環境との整合性が保持された全身動作の収録、または既存動作データの準備する。
- (2) (1)の動作データをアバタモデルに適用し対象の VR 環境において再生する。ユーザは実行映像を確認しながら、任意の身体部位でアバタを操作することを想定して動作を実行するとともに、その際のアバタの環境センサ情報を記録する。
- (3) (2)の目視からの想定する操作動作と対象

となる全身動作との時間遅れを補正し、(1)を教師データ、補正後の(2)を対応する入力データとする。

2.3. アバタ動作構成モデルの構築

本稿では、時系列データの時間的特徴を考慮できるリカレントニューラルネットワークを用い、Long Short-Term Memory (LSTM) ユニットによりモデルを構築した。モデルは入力層、出力層、LSTM層の計3層で構成され、入力ユニット数は上半身17関節の3軸の回転量の51および環境センサ情報の50の計101、出力ユニット数は全身51関節の3軸の回転量で153、LSTM層は1000とした。Backpropagation Through Time (BPTT)は30とした。ここで訓練データから得られる出力を期待される各関節回転量を \hat{p}_n 、モデルからの出力された各関節回転量を p_n とすると誤差関数Lは次式で表される。

$$L = \sum_n \|p_n - \hat{p}_n\|^2 \quad (1)$$

3. 実験

頭上に障害物があるVR環境におけるアバタの歩行動作を対象に実験を行った。

VR環境におけるアバタの対象動作は、頭上の障害物を避けるようにしゃがみながら歩行する動作および立位での歩行動作とし、既存のモーションキャプチャデータを使用した。また、操作を意図した座位姿勢の上半身による再現動作をKinect v2を用いて収録すると同時にVR環境における環境センサ値を取得し、以上を訓練データとした。総フレーム数は36,480フレーム(約20分)である。この訓練データを用いて2.3で定義したモデルの学習を行った。

学習済みモデルに対して、座位姿勢の上半身による動作および環境センサ値の入力からアバタ動作を構成した結果を図2, 3に示す。図2は頭上に障害物がない場合、図3は頭上に障害物がある場合のアバタの全身動作の構成結果である。

出力動作の精度の確認のため、入力動作は事前に用意した操作を意図した動作(正解動作)として、映像で確認しながら実行した。正解動作と出力動作の姿勢遷移より、上半身の動きに対応している意図した歩行動作が構成されることが確認できる。

また、図2, 3より類似した上半身における動作入力からでも環境センサの値の違いにより、立位での歩行動作、しゃがみ状態での歩行動作が構成できていることが確認できる。

4. おわりに

本稿では、ユーザの姿勢入力とCGアバタが置かれているVR環境における周囲の環境情報を基

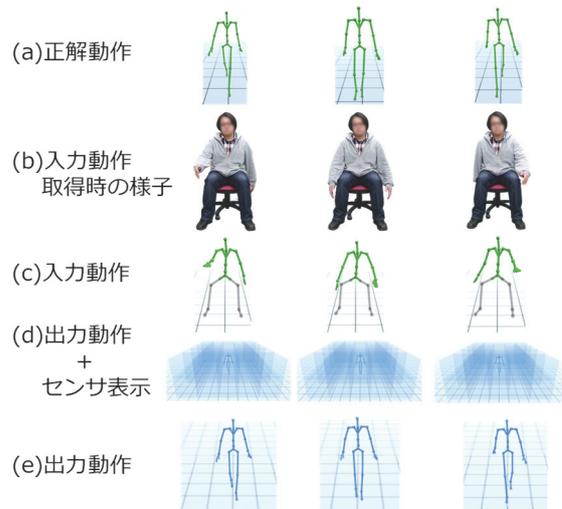


図2. 歩行動作の構成結果(障害物なし)

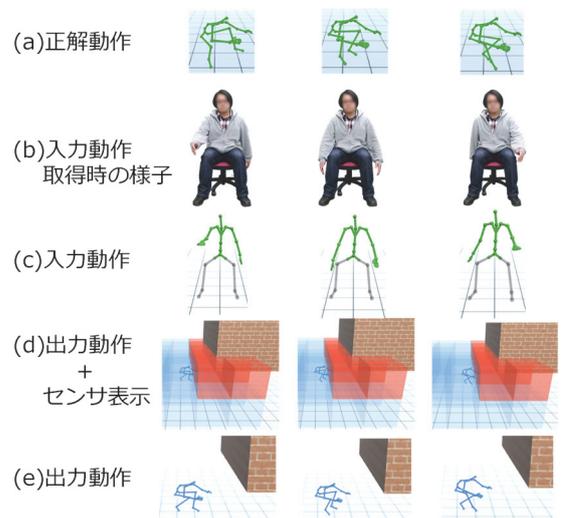


図3. 歩行動作の構成結果(障害物あり)

に、ユーザの意図を反映し且つVR環境と整合性のあるアバタ動作を構成する手法を提案した。歩行動作を対象とした実験結果より座位姿勢の上半身による入力からユーザの操作意図を反映し、且つ環境によって異なる全身の動作の構成を確認した。今後の課題として、様々な動作と環境を対象にした動作構成結果の検証とユーザの主観評価による操作意図の反映精度の検証が挙げられる。

参考文献

- [1] 仲泰誠他：“部分的な姿勢入力からユーザの操作意図を反映したアバタの全身動作の構成”，情報処理学会第82回全国大会第1分冊，pp. 401-402, 2020.
- [2] S. STARKE, H. ZHANG, T. KOMURA, J. SAITO：“Neural State Machine for Character-Scene Interactions”，ACM Transactions on Graphics, 38(6), Article No.178, 2019