



HMD センシング情報を用いた操作意図推定による アバタの動作構成

Using HMD Sensing Information to Estimate Operational Intent
Avatar Operation Configuration

西田有吾¹⁾, 森博志¹⁾, 外山史¹⁾

Yugo NISHIDA, Hiroshi MORI, and Fubito TOYAMA

1) 宇都宮大学 大学院地域創生科学研究科 (〒321-0904 栃木県宇都宮市陽東 7-1-2, y.nishida@vc.is.utusnomiya-u.ac.jp)

概要: 本稿では, VR-HMD 付属のセンサから取得可能な頭と手のトラッキング情報を用いてユーザの意図したアバタの全身動作を構成する手法を提案する. 深層学習を用いてユーザの直感的な操作入力情報から意図する全身姿勢情報を推定する. 実験では, 訓練データ作成者および非作成者を対象に座位姿勢における直感的な姿勢入力を用いて, アバタの全身動作を構成した. 目標姿勢との姿勢差の比較より, 意図した全身動作の構成が可能であることを確認した.

キーワード: ユーザインタフェース, アバタ, モーショントラッキング

1. はじめに

VR においてユーザの代替映像として用いられるアバタのフルボディトラッキングによる操作では, ユーザの姿勢がそのままアバタに反映されるため, 直感的な操作が可能である. 一方, ユーザは意図したアバタの動作と同一の動作を継続的に入力する必要があり, 操作に対する身体的・心理的な負担は大きい. この問題に対し, 座位姿勢による部分的な身体姿勢入力からのアバタ動作の構成手法 [1] では, 上半身の姿勢のトラッキング情報から意図したアバタの全身動作の構成を可能にしている.

また, VR-HMD 付属のセンサの充実により, 頭や手の位置・姿勢のトラッキング情報の利用が一般化しており, センサ取得情報からユーザの全身姿勢を推定する手法 [2] が提案されている. しかし全身姿勢の推定には同等の姿勢をとることが前提である.

そこで本稿では, VR-HMD 付属のセンサから取得可能な頭と手のトラッキング情報を用いてユーザの意図したアバタの全身動作を構成する手法を提案する. 本手法では, アバタの姿勢との一致を必要としないユーザの任意の安定姿勢における直感的な姿勢のトラッキング情報からユーザの意図するアバタの全身動作を構成する. そのため, 操作に対する身体的・心理的な負担の軽減による利便性の向上が期待できる.

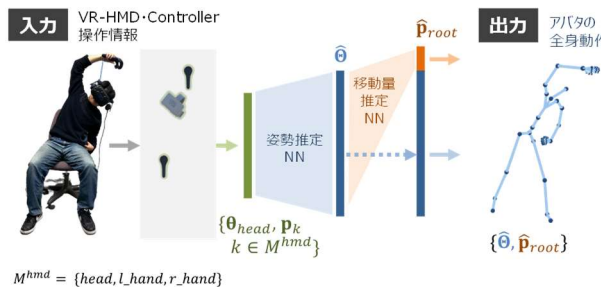


図 1 提案手法の概要

2. 提案手法

2.1 概要

提案手法の概要を図 1 に示す. 本手法では VR-HMD および付属のハンドコントローラから取得した頭と手の姿勢情報からユーザの意図するアバタの全身動作を構成する. 入力は頭部の姿勢 θ_{head} と位置 \mathbf{p}_{head} , および両手の位置 \mathbf{p}_{Lhand} , \mathbf{p}_{Rhand} で, 姿勢推定ネットワークによりアバタの全身姿勢 $\hat{\theta} = \{\hat{\theta}_i \mid i = hip, spine, \dots, head\}$ を推定する. 次に $\hat{\theta}$ より姿勢変化に伴うアバタの骨格モデルのルート位置 $\hat{\mathbf{p}}_{root}$ を推定する. 以上の処理を時系列で実行することによりアバタの全身動作 $\{\hat{\theta}(t), \hat{\mathbf{p}}_{root}(t)\}$ を得る.

2.2 ユーザの意図したアバタの全身姿勢の構成

各ネットワークの構築には, 時系列データの時間的特徴を考慮可能な RNN の構成手法の一つである LSTM を用いた. 姿勢推定ネットワークは入力層, LSTM 層, 出力層で構成され, 入力ユニット数は頭部関節の 3 軸回転量と各 3

次元位置, 3 関節の速度と加速度を合わせた 30, 出力ユニット数は全身 25 関節の 3 軸回転量に対応する 75 である.

移動量推定ネットワークの入力ユニット数は姿勢推定ネットワークから出力される全身 25 関節の 3 軸回転量に対応する 75, 出力ユニット数はルート関節の 3 軸の移動量に対応する 3 である. それぞれのモデルにおいて LSTM 層は 1000 とし, 損失関数には平均二乗誤差を用いた.

訓練データセットは, 操作者が正解動作を目視で確認しながら, その操作を想定して座位姿勢で直感的に動作を行い, HMD とハンドコントローラのトラッキング情報を記録することで作成する. 姿勢推定ネットワークの教師データは正解動作の全身姿勢情報, 入力データは操作者による入力であり, 移動量推定ネットワークの教師データは正解動作のアバタの移動量, 入力データは正解動作の全身姿勢情報である.

3. 実験

3.1 実験条件

上半身と下半身が連動する左右伸脚運動, 両手を上げる動作をはじめとする左, 右, 上方向それぞれで類似する動作を含む 11 種類の全身動作を対象に, 3 名の操作者に目標動作の操作を意図して, 座位姿勢で直感的な動作を行ってもらうことで訓練データを作成した. 訓練データの総フレーム数は 168408 フレーム (約 94 分) である.

出力動作の精度の検証のため, 操作入力時には正解動作を目視で確認しながらその動作を意図して直感的な操作入力を行うよう被験者に指示した. 訓練データの有無による構成精度の比較のため, 被験者は訓練データ作成者 3 名, および訓練データ非作成者 6 名の計 9 名である.

3.2 実験結果

図 2, 3 に浅い伸脚運動 (伸脚 1) と深い伸脚運動 (伸脚 2) のアバタ動作の構成結果を示す. 訓練データ作成者の構成結果 (図 2) では, 伸脚 1, 伸脚 2 とともに正解動作に近い動作が構成され, 両者の動作が適切に構成できていることが確認できる. また, 非作成者の構成結果 (図 3) の伸脚 2 では正解動作に近い動作が構成されていることが確認できる. 一方, 伸脚 1 (図 3(i)) の (c) (d) を比較すると, 正解姿勢と出力姿勢の差が大きいため確認できる.

表 1 に 11 種の全身動作を対象とした提案手法による構成結果と正解姿勢との平均関節位置誤差 (MPJPE) を示す. 訓練データ非作成者の構成結果の姿勢差は, 訓練データ作成者の構成結果と比較して, 誤差が大きいため確認できる. これは操作者により目標となる全身姿勢を意図する直感的な操作入力姿勢が異なる点, および自身の操作姿勢情報を用いた訓練データによる推論モデルの学習が行われておらず正確な全身姿勢が構成されなかったことが原因だと考えられる.

4. むすび

本稿では, VR-HMD 付属のセンサから取得可能な頭と

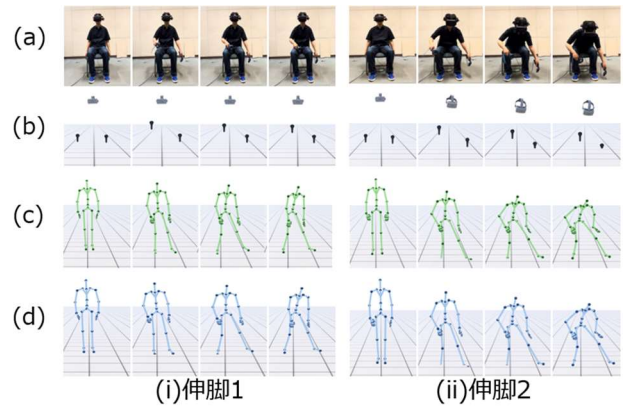


図 2 訓練データ作成者によるアバタ動作の構成結果.

(a) 入力の様子, (b) 入力姿勢情報, (c) 正解姿勢, (d) 出力姿勢.

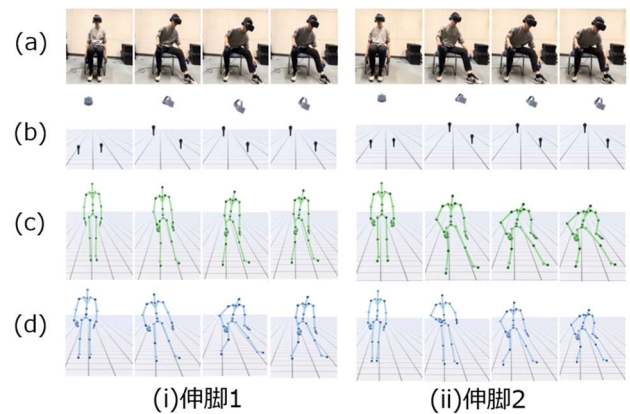


図 3 訓練データ非作成者によるアバタ動作の構成結果.

(a) 入力の様子, (b) 入力姿勢情報, (c) 正解姿勢, (d) 出力姿勢.

表 1: 11 種類の動作の平均関節位置誤差 (MPJPE)

	訓練データ作成者	訓練データ非作成者
姿勢差(cm)	4.668	10.152

手のトラッキング情報を用いたユーザの意図するアバタの全身動作を構成する手法を提案した. 実験より訓練データ作成者は高い精度で意図したアバタの全身動作を構成できたのに対し, 訓練データ非作成者の構成結果では, 類似の全身動作の判別が難しく, 誤差が大きくなる傾向を確認した.

今後の課題として, 訓練データの拡充による構成精度の向上と, アバタ周囲の環境情報を考慮した構成手法の検討が挙げられる.

参考文献

[1] K.Yamada et al, "Towards Controlling Whole Body Avatars with Partial Body-Tracking and Environmental Information," IEEE VRW 2022, pp.818-819, 2022.

[2] L.Suimin et al, "QuestEnvSim : Environment-Aware Simulated Motion Tracking from Sparse Sensors," ACM SIGGRAPH 2023 Conference Proceeding, Article:No62, pp.1-9, 2023.