

## コンテンツ整合性に基づく姿勢置換によるアバタの動作構成

稲葉 元哉<sup>†1</sup> 福田 冨斗<sup>†2</sup> 森 博志<sup>†3</sup> 外山 史<sup>†4</sup><sup>†</sup>宇都宮大学 大学院地域創生科学研究科

## 1 はじめに

ユーザの代替映像として用いられるアバタは、エンタテインメントやソーシャルコミュニケーションを対象とした VR コンテンツを中心に利用が進んでいる。近年では、設置型ディスプレイに表示されたアバタをホテルの受付やコンビニエンスストアにおける接客への活用のための試験運用が取り込まれ始めている。

アバタの操作方法として、ユーザの姿勢を計測しアバタの姿勢に反映するモーショントラッキング方式がある。ユーザの姿勢をアバタの姿勢にそのまま反映するため、ユーザは直感的なアバタの操作が可能になる。一方、ユーザが操作中に突発的に不適当な姿勢をとってしまった場合には、本来対面者に提示したくないアバタの姿勢も構成されてしまう。

部分的な姿勢置換によるアバタの動作補正手法[1]では、取得したユーザの入力姿勢情報に対し、骨格モデルの構造や身体運動の原理に基づいて信頼度を推定し、低信頼度部位の姿勢を部分的に置換することで対面者から自然に見えるアバタ動作を構成している。この手法では本来とり得ない破綻姿勢の推定と置換は可能であるものの、接客アバタを操作している際に客である対面者に見せたくない姿勢、例えば鼻をかく、脚を交差させるなど、生理現象によりやむを得ず取りうる不適当な姿勢や、無意識的にとる不適当な姿勢については、検出することが難しい。

そこで本稿では、利用用途において不適当な入力姿勢を検出し、適切な類似姿勢で部分的に置換することで、コンテンツ適合性を保持したアバタ動作を構成する手法を提案する。本手法により直観的な操作性を保持したまま対面者に適切なアバタ映像を提示可能になるため、ユーザ自身の姿勢を反映するモーショントラッキングによるアバタ操作の利用に対する心理的なハードルを下げる効果が期待できる。

## 2 コンテンツ整合性に基づくアバタの姿勢置換

## 2.1 概要

提案手法の概要を図1に示す。入力としてモ

Avatar Motion Synthesis by Joint Pose Replacement Based on Content Suitability

Motoya Inaba<sup>1</sup>, Sayato Fukuda<sup>2</sup>, Hiroshi Mori<sup>3</sup>, Fubito Toyama<sup>4</sup>

<sup>†</sup>Utsunomiya University

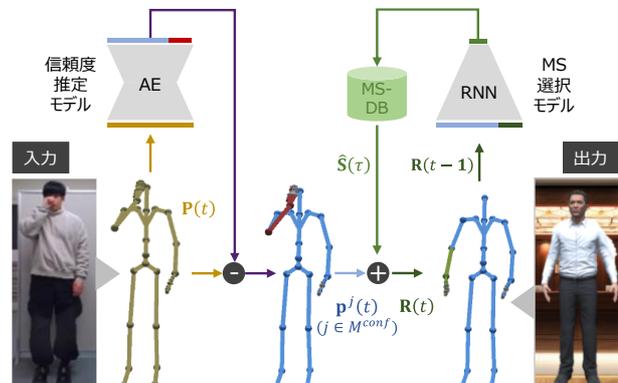


図1 提案手法の概要

ーショントラッキングによりユーザの姿勢情報である関節角回転量 $\mathbf{P}(t) = \{\mathbf{p}^j(t) | j \in M^{input}\}$ 、およびルート関節の位置情報 $\mathbf{z}(t)$ を取得する。 $t$ は実行時のフレーム番号、 $j$ は関節部位、 $M^{input}$ は入力姿勢の関節部位の集合を表す。次に、取得した姿勢情報 $\mathbf{P}(t)$ について関節ごとに信頼度を推定する。Autoencoderを用いて構築した信頼度推定モデルにより、適切な姿勢と推定される関節部位を高信頼度部位 $M^{conf}$ 、適切な姿勢ではないと推定される関節部位を低信頼度部位 $\overline{M^{conf}}$ とする。最後に、前フレームの出力姿勢 $\mathbf{R}(t-1) = \{\mathbf{r}^j(t-1) | j \in M^{output}\}$ を、モーションセグメント選択モデルに入力し、ユーザの姿勢の置換に用いるモーションセグメント $\hat{\mathbf{S}} = \{\hat{\mathbf{s}}^j(\tau) | j \in M^{output}\}$ を蓄積動作群から選択する。 $M^{output}$ は出力姿勢の関節部位の集合である。選択されたモーションセグメントにより、ユーザの入力姿勢の低信頼度部位を置換し、高信頼度部位は入力姿勢をそのまま反映することで、出力姿勢 $\mathbf{R}(t) = \{\mathbf{r}^j(t) | j \in M^{output}\}$ を構成する(式1)。

$$\mathbf{r}^j(t) = \begin{cases} \mathbf{p}^j(t) & (j \in M^{conf}) \\ \hat{\mathbf{s}}^j(\tau) & (j \in \overline{M^{conf}}) \end{cases} \quad (1)$$

## 2.2 コンテンツ不適合姿勢の推定

アバタの出力姿勢として対面者に提示するには不適当な姿勢の推定には、Autoencoderを使用した異常検知[2]を用いる。本手法における信頼度推定モデルは入力層のユニット数を25関節3軸の回転量の75、中間層を64-32-64、出力層を入力層と同様75としたLSTM-Autoencoderにより構築した。損失関数は平均二乗誤差とし、考

慮する時系列区間は 30fps で取得した 30 フレームとした。

### 2.3 類似姿勢による不適合姿勢の置換処理

ユーザの前フレームの出力姿勢 $\mathbf{R}(t-1)$ を入力として LSTM を用いて構築したモーションセグメント選択モデルにより選択された類似モーションセグメントの各姿勢 $\hat{\mathbf{S}}(\tau)$ について、各関節回転量に基づく姿勢距離が最小となるフレーム $\tau^{similar}$ を次式により算出する。

$$\tau^{similar} = \arg \min_{0 \leq \tau \leq T} \sum_j |\mathbf{r}^j(t-1) - \hat{\mathbf{s}}^j(\tau)| \quad (2)$$

算出したフレームの姿勢 $\hat{\mathbf{S}}(\tau^{similar})$ は前フレームの出力姿勢との姿勢距離が最小となる姿勢のため、次フレームの $\tau^{similar} + 1$ を開始フレーム $\tau^{connect}$ とした類似モーションセグメント $\hat{\mathbf{S}}(\tau) = \{\hat{\mathbf{s}}^j(\tau) | j \in M^{output}, \tau^{connect} \leq \tau \leq T\}$ により低信頼度部位の姿勢を置換する。

## 3 実験

### 3.1 実験条件

本実験では、アバタによる接客コンテンツを想定した動作構成結果を確認する。

信頼度推定モデルの学習には、対面コミュニケーションを対象とした動作を 30fps で取得した計 34,692 フレームの動作データを訓練データとして用いた。モーションセグメント選択モデルの学習には、上記動作データの内、7805 フレームの動作データを姿勢の類似度に基づいて分割した 27 個のモーションセグメントを用いた。モーショントラッキングデバイスには Microsoft Azure Kinect を用いた。

### 3.2 実験結果

鼻をかく姿勢を含む入力動作に対する本手法の適用結果を図2に示す。図2(a)の左手で鼻をかくような不適切な姿勢が低信頼度姿勢として推定され、類似姿勢で部分的に置換された結果、直立姿勢が構成されていることが確認できる。

信頼度推定モデルの評価として、鼻をかく動作を対象とした再現率・適合率・正解率を表1に、ROC 曲線を図3に示す。人の判断基準にて適切・不適切のラベル付けにより得た正解データと、提案手法による推定結果とを比較した。AUC の値は 0.92 であり、提案手法は高い精度で想定外の不適切な姿勢を推定可能であることが確認できる。

## 4 おわりに

本稿では、特定のコンテンツにおいて不適切な姿勢を隠匿し対面者に適切なアバタの姿勢を提示するためのアバタの動作構成手法を提案し

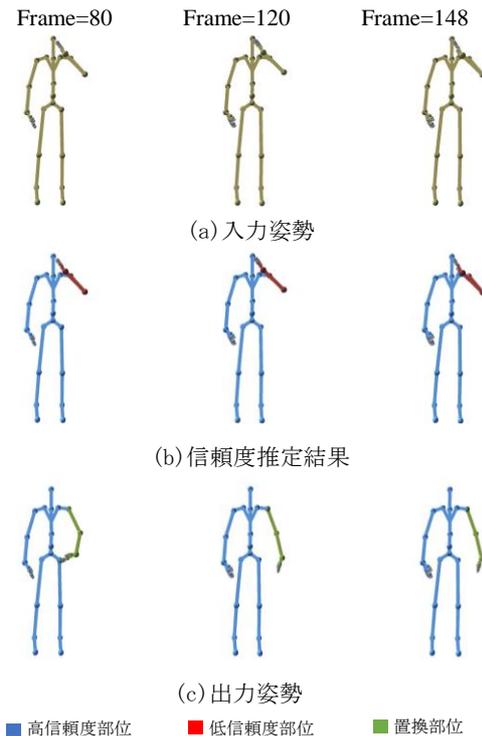


図2 不適合姿勢の推定と置換結果

表1 再現率・適合率・正解率

	再現率	適合率	正解率
従来手法 [1]	0	0	0.54
提案手法	0.83	0.96	0.89

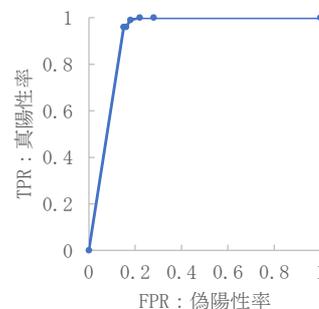


図3 ROC 曲線

た。実験では、アバタによる接客コンテンツを想定した対面コミュニケーション動作を対象に、隠匿したい不適切な姿勢をとった際、高い精度で不適切な姿勢をとる関節部位を検出可能であることを確認した。今後の課題として、動作補正結果の適合性の評価が挙げられる。

### 参考文献

- [1] 森他, “低信頼度の姿勢情報の置換により姿勢破綻を軽減した CG アバタ動作の実時間構成,” 日本バーチャリアリティ学会, vol.22, no.1, pp.19-26, 2017.
- [2] Y.Weï et al., “LSTM-Autoencoder based Anomaly Detection for Indoor Air Quality Time Series Data,” IEEE Sensors Journal, vol.23, no.4, pp.3787-3800, 2023.