

高信頼度姿勢情報と蓄積モーションセグメントの合成によるCGアバタ動作の構成 Synthesizing Avatar Motion from Joint Poses and Motion Segments

田中 里奈[†] 森 博志[†] 外山 史[†] 東海林 健二[†]
Rina Tanaka Hiroshi Mori Fubito Toyama Kenji Shoji

1. はじめに

Kinectをはじめとした深度センサを利用したモーションキャプチャでは、マーカ等を身につける必要がないため気軽にユーザの動きに応じたCGアバタの操作が可能である。

しかし、測定誤差等の影響により姿勢情報の一部が正確に取得できず、その情報を反映すると不自然に見えるアバタ動作が生成される場合がある。この問題に対し、深度センサを利用したモーションキャプチャの精度向上を目的とした、全身動作の正確な推定法が提案されている[1]。

一方、ユーザの動きに対応したアバタの動作アニメーションを実現する際、入力動作と完全に一致していても出力されるアバタの動作アニメーションが自然に見えれば要求を満たす場合が考えられる。そこで本稿では、取得姿勢情報のうち、正確に取得できている姿勢情報を抽出し、その情報を基に蓄積モーションデータの中から選出した最適なモーションセグメントで補填し合成することで、出力として自然に見えるアバタ動作を構成する手法を提案する。

2. 提案手法

2.1 手法概要

手法の概要を図1に示す。事前処理として、モーションキャプチャを用いてモーションデータを収録し姿勢差を基にセグメントに分割する。

アバタ操作の実行時処理は次の手順で行う。まず、深度センサを用いて姿勢情報として各関節部位の角度情報とルート部位の位置情報を取得する。次に、取得した各関節部位の角度情報に対して信頼度を算出し、高信頼度部位を推定する。最後に、推定した高信頼度部位を基に蓄積モーションセグメントから最適なセグメントを選出し、低信頼度部位と未取得部位に補填し合成することで、自然に見えるアバタ動作を構成する。

2.2 入力姿勢情報からの高信頼度部位情報の抽出

全身の入力姿勢情報から正確に取得できていると推定される関節部位の姿勢情報を抽出する。

使用するセンサによってはAPIから各関節部位のトラッキング状態を取得できるが、トラッキング状態が真の場合でも動作が自然に見えない場合がある。

そこで本手法ではトラッキング状態の情報に加え、従来の取り組み[2]と同様に入力姿勢情報に対して信頼度を算出することで、正確に取得できていると推定される関節部位を抽出する。入力姿勢情報の各関節部位の角度情報を利用して、蓄積モーションデータに基づいた姿勢遷移の連続性と姿勢の拘束性、および関節部位間の干渉と階層構造の親

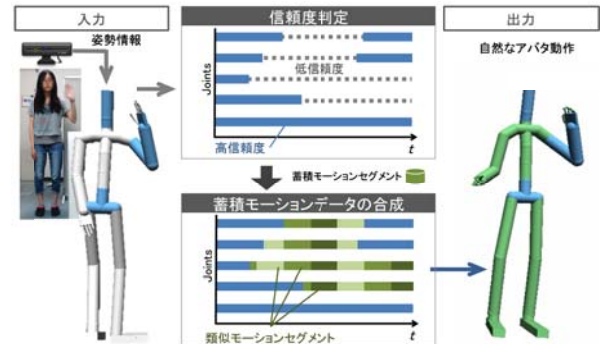


図1 提案手法

子関係に基づいて信頼度を求める。

得られた信頼度に基づいてあるフレーム t の入力姿勢情報 $\mathbf{P}(t) = \{\mathbf{p}^j(t) \mid j \in M^{input}\}$ の各部位 j を、正確に取得できていると推定される高信頼度部位 M^{conf} 、正確に取得できていないと推定される低信頼度部位 $\overline{M^{conf}}$ に分類する。また、出力対象のアバタに含まれる関節部位 M^{output} の中で入力姿勢としてセンサで取得できない部位を未取得部位 M^{na} とする。

以上より、あるフレーム t の入力姿勢を $\mathbf{R}(t) = \{\mathbf{r}^j(t) \mid j \in M^{output}\}$ とすると、 $\mathbf{r}^j(t)$ は次式のように表すことができる。

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

なお、 $\mathbf{s}_n^j(t)$ はモーションセグメント n の関節 j の角度情報である。

2.3 蓄積モーションデータのセグメント化

低信頼度部位や未取得部位の姿勢情報の補填に利用する蓄積モーションデータを、本手法では連続した姿勢類似度により小単位に分割したセグメントで扱う。

モーションセグメントは、蓄積モーションデータ $\mathbf{Q}(t) = \{\mathbf{q}^j(t) \mid j \in M^{output}, 0 < t < T^{sample}\}$ を姿勢の類似度を基に分割することで構成し、セグメント $\mathbf{S}_n(t)$ を $\mathbf{S}_n(t) = \{\mathbf{s}_n^j(t) \mid j \in M^{output}, 0 < t < T^{S_n}\}$ と表現する。ただし、 $0 \leq n < N$ である。

モーションセグメントの構成について、動作速度を基に分割する手法[3]等が提案されている。本手法では見た目の姿勢差を考慮し、全身の関節部位の位置情報による姿勢差を用いて構成する。蓄積モーションデータ $\mathbf{Q}(t)$ に対応する各関節部位の位置を $\mathbf{G}(t) = \{\mathbf{g}^j(t) \mid j \in M^{output}, 0 < t < T^{sample}\}$ として、次の手順でセグメントを構成する。

[†] 宇都宮大学 Utsunomiya University

- (1) $n = 0$ とし, $\mathbf{Q}(t)$ の先頭フレームを起点 t_{s_n} とする.
- (2) $t_{s_n} < h \leq T^{sample}$ を満たすフレーム h において起点フレームの姿勢 $\mathbf{G}(t_{s_n})$ との姿勢距離 Δg_h を次式で求める.

$$\Delta g_h = \sum_j |\mathbf{g}^j(t_{s_n}) - \mathbf{g}^j(h)| \quad (2)$$

- (3) Δg_h が閾値 $g^{threshold}$ を超えた場合, $\mathbf{Q}(t) = \{\mathbf{q}^j(t) \mid j \in M^{output}, t_{s_n} < t \leq h\}$ をセグメント $\mathbf{S}_n(t) = \{\mathbf{s}_n^j(t) \mid j \in M^{output}, 0 < t < T^{s_n}\}$ とする. ただし, $T^{s_n} = h - t_{s_n}$ である.
- (4) $t_{s_n} = h, n = n + 1$ とし手順(2)に戻り処理を続けセグメントに分割する.

また, 取得したセグメント n の総フレーム数 T^{s_n} が閾値 $T^{threshold}$ 以上の場合, 下の式を満たすフレーム t_d を基にセグメントを細分化する.

$$t_d = \arg \max_t \sum_j |\mathbf{g}^j(t) - \mathbf{g}^j(t-1)| \quad (3)$$

以上の処理により, サンプルデータ $\mathbf{Q}(t)$ から T^{s_n} フレーム以下の長さを有する N 個のモーションセグメントを得る.

2.4 高信頼度部位情報に基づいた蓄積モーションセグメントの合成

実行フレーム t^{now} において $j \in M^{conf}$ の組み合わせが変化, あるいは $t^{now} - 1$ フレームにおいてモーションセグメントの最終フレームが補填されていた場合, 次の手順で t^{now} において補填するモーションセグメントを選択し合成する.

- (1) フレーム t^{now} における高信頼度部位の姿勢情報 $\{\mathbf{p}^j(t^{now}) \mid j \in M^{conf}\}$ と各モーションセグメントの高信頼度部位に該当する関節の姿勢情報の時間平均を比較し, 補填候補のモーションセグメントを選択する.
- (2) 補填候補のモーションセグメントにおいて, フレーム t^{now} で最も滑らかに接続されるセグメント $\mathbf{S}_{connect} = \{\mathbf{s}_{connect}^j(t) \mid j \in M^{output}, 0 < t < T^{s_{connect}}\}$ とセグメント $\mathbf{S}_{connect}$ 中のフレーム $f^{connect}$ を求める.
- (3) $\mathbf{r}^j(t^{now} - 1)$ と $\{\mathbf{p}^j(t^{now}) \mid j \in M^{conf}\}$, $\{\mathbf{s}_{connect}^j(f^{connect}) \mid j \in M^{conf}, M^{na}\}$ が滑らかに接続するように姿勢情報を合成する.

次項より各手順について説明する.

2.4.1 補填モーションセグメント候補の選出

フレーム t^{now} における高信頼度部位の姿勢情報 $\{\mathbf{p}^j(t^{now}) \mid j \in M^{conf}\}$ と, 各モーションセグメントの低信頼度部位および未取得部位に該当する関節の姿勢情報 $\{\mathbf{s}_n^j(t) \mid j \in M^{conf}, M^{na}, 0 < t < T^{s_n}\}$ の時間平均 \mathbf{s}_n^j を比較し, 補填候補のモーションセグメントを選択する.

$\mathbf{p}^j(t^{now})$ と \mathbf{s}_n^j の関節回転量による姿勢距離を基に, k 近傍法を用いて K 個の補填モーションセグメント候補 $\widehat{\mathbf{S}}_k(t) = \{\widehat{\mathbf{s}}_k^j(t) \mid j \in M^{output}, 0 < t < T^{s_k}\}$ を選出する.

2.4.2 補填モーションフレームの決定

補填候補の K 個のモーションセグメント $\widehat{\mathbf{S}}_k(t)$ において, 実行フレーム t^{now} に最も滑らかに接続されるフレーム \widehat{f}_k を求め, 各 \widehat{f}_k で接続する際に最も滑らかに接続されるセグメントを $\mathbf{S}_{connect}$ とする. $\mathbf{S}_{connect}$ は,

$\mathbf{S}_{connect} = \arg \min_{\widehat{\mathbf{S}}_k} \sum_j \sum_{l=1}^d w^j |\mathbf{r}^j(t-l) - \widehat{\mathbf{s}}_k^j(f-l)| \quad (4)$ として求められる. $\mathbf{S}_{connect}$ の接続フレーム \widehat{f}_k を $f^{connect}$ とする.

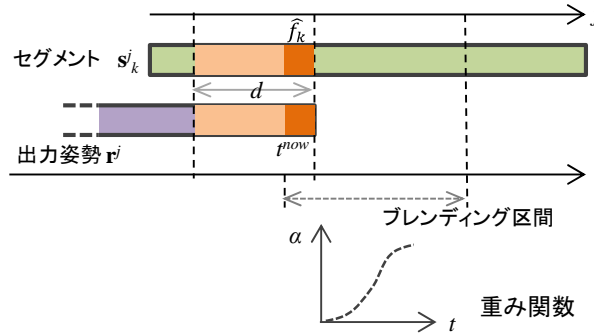


図2 モーションセグメントと姿勢情報の比較と合成

2.4.3 蓄積モーションデータの合成

2.4.2で求めた補填モーションフレーム $\mathbf{S}_{connect}(f^{connect})$ をフレーム t^{now} において接続する. その際, $\mathbf{r}^j(t^{now} - 1)$ と $\{\mathbf{p}^j(t^{now}) \mid j \in M^{conf}\}$, $\{\mathbf{s}_{connect}^j(f^{connect}) \mid j \in M^{conf}, M^{na}\}$ が滑らかな姿勢遷移となるように重み関数を基にブレンドする(図2).

3. 実験

3.1 実験概要

本実験では, 深度センサに Kinect を利用し, その API として OpenNI と NiTE を用いて最大で 15 関節の姿勢情報を 30fps で取得した. 蓄積モーションデータには事前にモーションキャプチャにより取得した, 直立姿勢を基準とし両腕を軽く動かす動作を利用した.

3.2 実験

3.2.1 遮蔽物の影響による測定誤差を含む入力動作を対象としたアバタ動作の構成

センサとユーザの間に遮蔽物がある環境下において, 直立姿勢を基準とした右手を振る動作を入力姿勢情報に提案手法を適用したときの各関節へのモーションセグメントの補填結果を図3にアバタ動作の生成結果を図4に示す.

図3より高信頼度部位の抽出とセグメントの補填が行われていることが確認できる.

図4より, 取得姿勢において右腕が腰と干渉しているため, 右腕が低信頼度部位と推定されていることが確認できる. また, 左腕も遮蔽物の後方を通過する際に姿勢情報を正しく取得することができず, 低信頼度部位と推定されている. 両脚は測定誤差により大腿部から後方を向いており, 低信頼度部位と推定されていることが分かる.

背骨や手指, 足指は本センサで取得することができないが出力対象のアバタに含まれる関節部位であり, 未取得部位となっている.

これらの低信頼度部位, 未取得部位を, 高信頼度部位の姿勢情報を基に蓄積モーションセグメントから選択したセグメントで補填した結果, 自然に見えるアバタ動作を構成できていることが確認できる.

3.2.2 選択された任意の関節部位の入力動作を対象としたアバタ動作の構成

着座姿勢で右手を振る動作を取得し, 右腕と首, 頭部を選択して入力姿勢情報とし提案手法を適用した結果を図5に示す.

図5より、入力として与えた右腕が、測定誤差により不連続な動きをしている際に、低信頼度と推定されていることが確認できる。

入力として与えた関節部位の姿勢情報を基に選択した蓄積モーションセグメントで補填した結果、自然なアバタ動作を構成できていることが確認できる。

低信頼度と推定されていた右腕の不自然な動作も、その前後の高信頼度部位として抽出されている動作と滑らかに接続されており、自然な動作が構成できている。

このように任意の関節部位を選択して入力姿勢として与えることにより、該当の関節部位のみから全身姿勢を構成することが可能であることが確認できた。

4. おわりに

本稿では、正確に取得できている姿勢情報を抽出し、その情報を基に蓄積モーションデータの中から選出した最適なモーションセグメントで補填し合成することで、出力として自然に見えるアバタ動作を構成する手法を提案した。

実験の結果、高信頼度部位を基に自然な全身のアバタ動作を構成することができた。本実験において構成したアバタ動作はいずれも30fpsで描画できており、本手法は実時間で利用することが可能であると考えられる。

また、任意の関節部位を選択して入力姿勢として与えることで、該当の関節部位のみから全身姿勢を構成できることが確認できた。これにより、例えば、会話時の頭部と腕部の動作から、直立姿勢を基準とした自身のアバタ動作を構成することができるため、ユーザの動作の一定のプライバシーを担保したアバタ映像を構成することが可能である。そのため本手法は3DCGアバタを利用したテレコミュニケーション等への応用が期待できる。

参考文献

- [1] Xiaolin Wei, Peizhao Zhang, Jinxiang Chai, "Accurate Realtime Full-body Motion Capture Using a Single Depth Camera," ACM Transactions on Graphics, Vol.31, Issue 6, No.188 (2012).
- [2] Rina Tanaka, Hiroshi Mori, Fubito Toyama, Kenji Shoji, "Real-time Avatar Motion Synthesis by Replacing Low Confidence Joint Poses," ACM SIGGRAPH2014 Posters, No.14 (2014).
- [3] Sophie Jörg, Jessica Hodgins, Alla Safonova, "Data-driven finger motion synthesis for gesturing characters", ACM Transactions on Graphics, Vol.31, Issue 6, No.189 (2012).

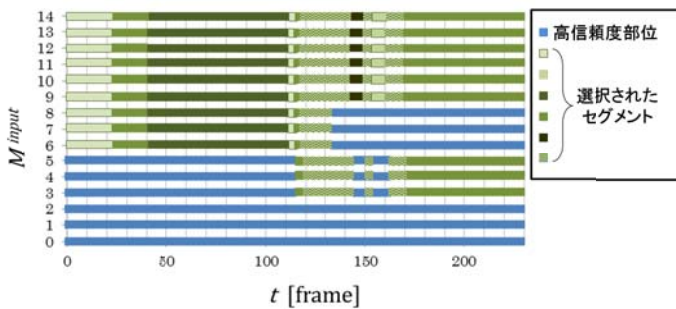


図3 蓄積モーションセグメントによる補填結果

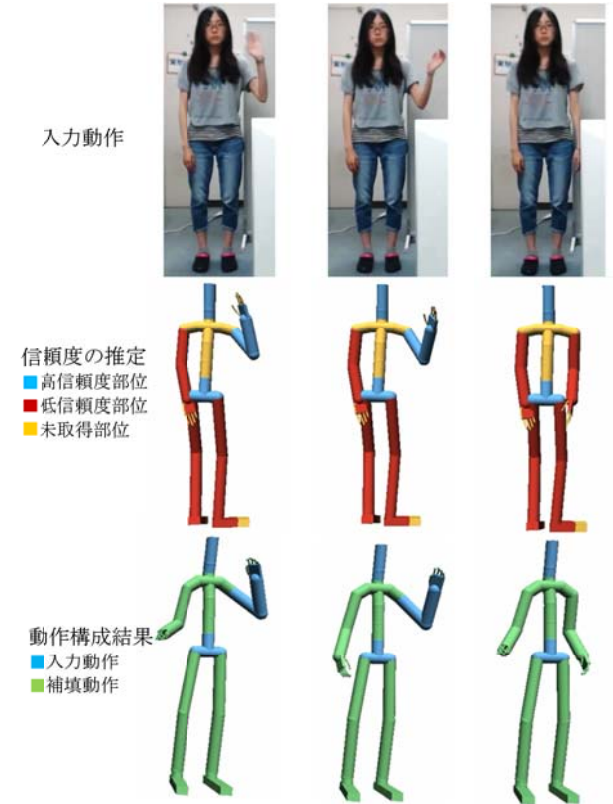


図4 遮蔽物の影響による測定誤差を含む入力動作を対象としたアバタ動作の構成結果

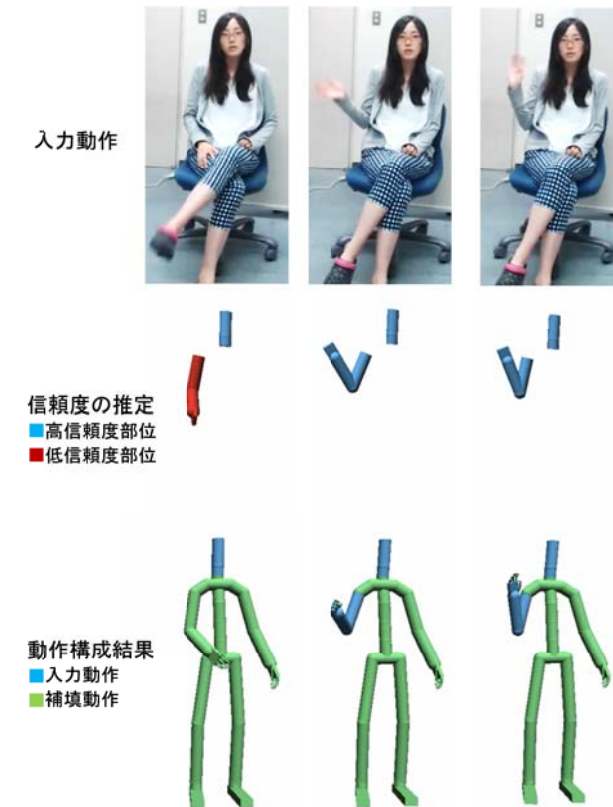


図5 選択された任意の関節部位の入力動作を対象としたアバタ動作の構成結果