

複数 Kinect ボーン情報を組み合わせた 高精度モーションキャプチャシステムの構築

吉本誠也[†] 高野茂[‡] 岡田義広^{††}

九州大学システム情報科学府[†]

九州大学大学院システム情報科学研究院[‡]

九州大学附属図書館付設教材開発センター^{††}

1. はじめに

近年三次元コンピュータグラフィックス (CG) の利用機会が増加している。三次元 CG はデジタルゲームや映画、シミュレーションなど多くの分野で利用されるが、それらの制作にはモーションデータが必要となる。モーションデータの作成にはモーションキャプチャ装置を利用する他、手作業で行う場合もある。いずれも実用レベルのモーションデータの生成には、高いコストが必要となる。

2010年、ジェスチャーや音声認識によって直観的なゲームプレイを可能にすることができるインタフェースデバイス Kinect が登場し、2011年に Kinect を Windows で利用するための SDK が公開されたことでモーションデータを比較的安価で簡便に取得することができるようになった。しかし単一の Kinect では正面以外を向いている場合の精度が悪く、不可視部位のトラッキングができないという問題がある。

本研究では複数台の Kinect を用いることによって上記問題の解決を図る。

2. Kinect

Kinect から得られる情報には RGB カメラの情報、赤外線カメラの深度情報、深度情報を基に出力されるボーン情報、マイクロホンのサウンド情報の4つがある。赤外線カメラから得られた情報と予め用意されている学習木から関節位置の予測を行い約 30FPS でリアルタイムにボーン情報 (関節位置) を出力することができる。しかし、演者が Kinect の方を向いている正面向きを前提として学習木が作られているため、演者が正面以外を向いている場合精度が落ちてしまう。

そこで、筆者らはボーン情報のみを複数の Kinect から取得し、それらを組み合わせることによって高精度なモーションデータを生成するシステムの構築を目指している。また、不可視部位が発生した場合の推定データを出力している状態 (Inferred) か測定データを出力している状態 (Tracked) かを示すパラメータ JointTrackingState も出力されており、本研究ではこのパラメータの値を出力データの信頼度として利用する。

3. 前準備

高精度のモーションデータを作成するために使用するデータとしてはボーン情報から得られる右手系の三次元座標に加え、以下で説明する演者の向きとトラッキング精度を利用する。

3.1 演者の向き

上記で説明したように、出力データの精度の高さを示す指標として演者の向きが重要であり、それを定義する必要がある。図1に示す赤い矢印を演者の向きであると定義する。

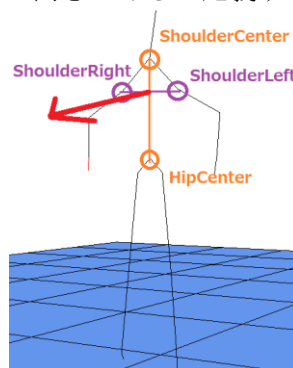


図1: 演者の向き

3.2 トラッキング精度

JointTrackingState が Tracked の数をトラッキング精度とする。JointTrackingState は不可視領域が多いほど Inferred の数が増えるため、複数の Kinect から出力されるポーズデータにつ

A motion capture system using multiple kinect skeleton tracking
[†]Seiya Yoshimoto, Graduate School of ISEE, Kyushu University
[‡]Sigeru Takano, Graduate School of ISEE, Kyushu University
^{††}Yoshihiro Okada, ICER of Kyushu University Library

いて不可視領域の少ない高精度のものを選択して使用する。

4. モーションデータの生成

複数 Kinect のモーションデータを組み合わせることで高い精度のモーションデータを生成する際、フレーム時間の同期、精度のよいポーズデータの選択、ポーズデータの補正処理が必要である。

4.1 フレーム時間の同期

フレーム時間の同期のために、図に示すようなキャリブレーション動作を導入する。まず初期動作として右手と左手を水平に伸ばしたポーズをとる。初期動作を確認できた PC からモーションデータを生成保存していく。演者はすべての PC が初期動作を認識できたことを確認した後でひじを曲げて振り子運動を開始する。振り子運動の周期を合わせる処理により同期を取ることが可能である。

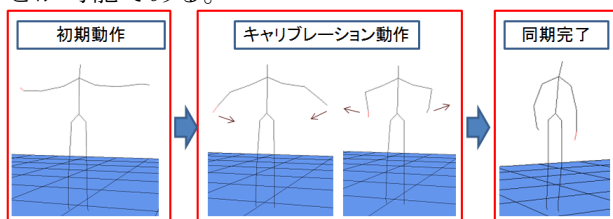


図 2: フレーム時間の同期

4.2 ポーズデータの選択

複数 Kinect のモーションデータについて、各フレームにおいてどの Kinect のポーズデータを使用すべきかを選択する。前節で定義した演者の向きとトラッキング精度を利用して、図 3 に示すように選択していく。まず、トラッキング精度が高いものを優先する。トラッキング精度が同じ高さの場合には、演者の向きが各 Kinect に対して正面に近いものを優先する。

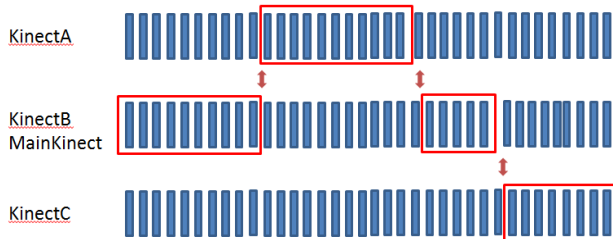


図 3: ポーズデータの選択

4.3 ポーズデータの補正

上記のように、各フレームにおいて、最も精度の高いポーズデータがそれぞれ選択されるが、異なる Kinect から出力されるデータは異なる 3 次元座標空間であるため、共通の 3 次元座標空

間のデータとなるよう補正処理を行う必要がある。キャリブレーションの初期状態で主となる Kinect を一つ選択し、その第一 Kinect の 3 次元座標空間のデータと合うように、他の Kinect から出力されるデータを補正する。図 4 に示すように第一 Kinect のポーズデータの赤色ボーンと紫色ボーンにそれぞれ一致するよう他の Kinect のポーズデータをアフィン変換する。

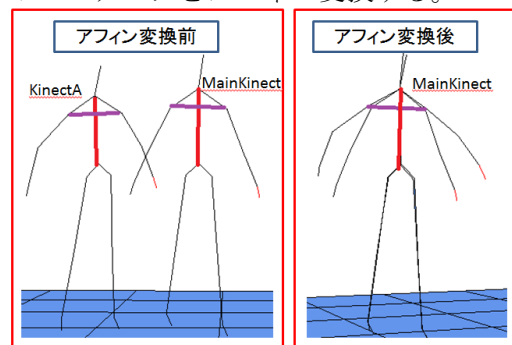


図 4: ポーズデータの補正

5. 精度の評価実験

フレーム時間とデータの精度を評価する実験を行う予定である。Kinect は 2 台使用し 2 台の PC にそれぞれ 1 台ずつ接続して使用する。まず、フレーム時間の同期の精度については、2 つのモーションデータの同期開始時刻が実際の同一フレームと何フレームずれているかを PC 時間の比較により計測する。次に、最終的に出力されるモーションデータの精度を評価する。評価としてはミス率を計算する。ミス率は各関節データについて前フレームの値と現フレームの値の差分をとり、それが一定値を超えた場合には、あり得ない位置に関節があると仮定し、それをミスとして扱う。全フレームに渡りミス率を計算する。

6. まとめ

今回複数の Kinect のデータから精度のよいモーションデータを簡便に作り出すシステムを提案した。今後の課題として、複数台の Kinect をどのように配置すると精度が上がるか、リアルタイム出力できないか等について検討する予定である。

参考文献

[1] Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio : Real-Time Human Pose Recognition in Parts from Single Depth Images