

## Abstract

全身運動を行う人物を一台のカメラで撮影し、三次元骨格推定を行う二種類の機械学習による方法を実装する。一方は RGB 画像を入力として用いる方法で、他方は RGB 画像に加え物体の奥行き情報 (デプス) を取得できるカメラを用いる方法である。慣性式モーションキャプチャデバイスによる計測結果と比較するためのキャリブレーション方法を検討し、定量的な比較を行った。慣性式モーションキャプチャデバイスによる計測結果では奥行き方向が縮んでいることがわかった。また、RGB 画像による方法では更に歪んでいることが読み取れ、連続した動作の測定には適さないことがわかった。

## 1 研究背景・目的

人の動きなどのノンバーバルな情報を、コミュニケーションに用いたり、デジタルアーカイブに利用したりするために、カメラの二次元画像から三次元の骨格情報を推測する技術が求められている。本研究室では柔道の三次元の動きを複数視点の動画から推測する研究<sup>[1]</sup>が行われた。しかし、この方法では複数台のカメラを設置し、キャリブレーションを行うため、十分な広さを持つ測定空間が必要である。また、近年物体までの奥行きの情報 (デプス) を取得できるカメラが市販されている。機械学習によって単眼カメラから深度推定を行う方法も提案されている<sup>[3]</sup>。

そこで本研究では機械学習を活用して一台の入力装置で三次元骨格推定を行う手法として、RGB 画像を入力する方法と RGB 画像に加えてデプス情報を入力する方法を用いて実装する。さらに慣性式モーションキャプチャデバイスによる計測結果と比較するための座標系及び時間のキャリブレーション方法を検討する。

## 2 研究内容

人の動作の三次元骨格推定を行う方法として、画像処理による方法やモーションセンサによる方法がある。画像処理による三次元骨格推定は撮影するカメラに、色情報を記録できる一般的な RGB カメラを用いる方法と、物体のデプスも取得可能な RGBD カメラを用いる方法がある (2.2)。

モーションセンサによる方法は、光学式や慣性式などがある。光学式は体表面にマーカーを取り付けそのマーカーを複数台のカメラで取り込むことで骨格を高精度に推定できるが広い計測空間が必要になる。慣性式は加速度、角速度、方位を測定できるセンサを体表面の指定箇所に取り付けることで骨格を推定する (2.1)。

本研究では、市販の入力デバイスを使用する 3 つの推定方法を実装して性能を比較評価する。

### 2.1 慣性式モーションキャプチャの三次元骨格推定

慣性式モーションキャプチャデバイスとして mocopi を用いる。mocopi のモーションデータのフレームレ

率は 60 fps である。センサは 3 つの自由度を持つ角度センサと加速度センサを搭載している。両手、両足、頭、腰の計 6 ヶ所にセンサを装着してリアルタイムに三次元計測を行うことができる。センサを装着しない肘や膝などの関節部を直接測定することはできないが、機械学習を用いることで図 2(a) に示すような 27 個の関節位置を推定している。

mocopi のモーションデータは BVH ファイルで出力される。

### 2.2 画像処理による三次元骨格推定

画像処理を用いて三次元骨格推定をしている様子を図 1 に示す。本研究では画像入力装置として RGBD カメラである Intel RealSense D415 (以下 RealSense) を使用する。カメラの解像度は最大 1280×720 px、フレームレートは 60 fps であり、デプスセンサの測定範囲は 50 cm から 3 m である。

画像処理による方法として 2 つの方法を実装する。

1 つ目はカラー画像を入力として Google が提供するオープンソースの機械学習ライブラリ MediaPipe Pose を用いることで図 2(b) に示す 33 個の関節の三次元骨格情報を取得する。

2 つ目はカラー画像とデプス情報を入力として 3DiVi Inc が提供するライブラリ NuiTrack を用いることで図 2(c) に示す 19 個の関節の三次元骨格情報を取得する。

### 2.3 キャリブレーション

推定方法によって骨格座標のスケールや基準となる座標系が違うので定量的に比較評価するためには座標系を統一する必要がある。そこで計測開始時に、両腕

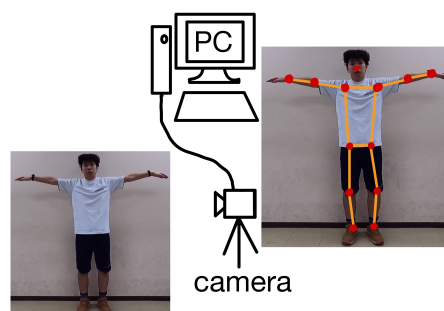
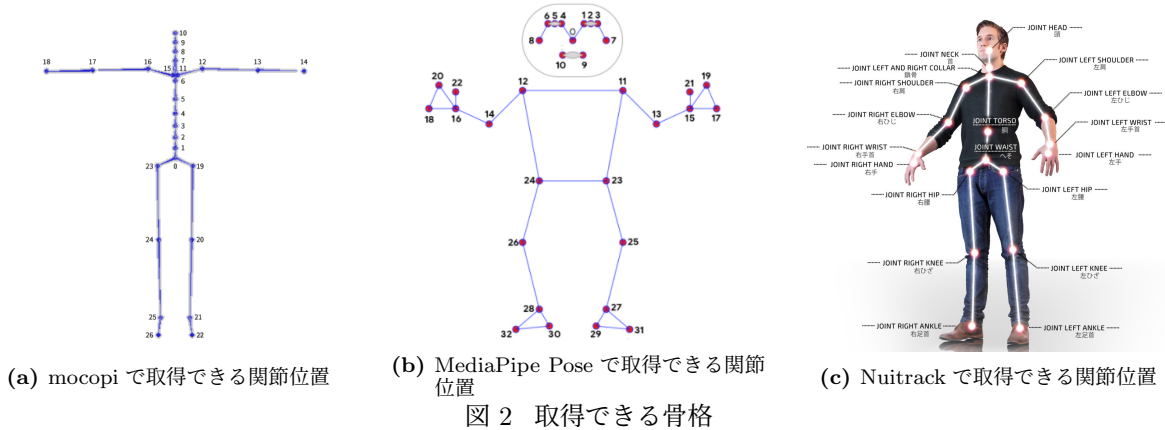


図 1 画像処理による骨格推定の様子



を水平に上げるポーズを取ることにする．水平に上げた両手首の距離を元にスケールを合わせる．また，へその位置を原点として，頭に向かう方向を  $y$  軸，右手から左手に向かう方向を  $x$  軸，これらに軸の直行する方向を  $z$  軸と定める．

動作の比較を行うには同期を取る必要があるが mocopi は同期計測ができない．そこで座標系を合わせるポーズの後で両手を伸ばしたまま胸の前で合わせるポーズをする．手が合わさっている時，両手首が最接近するので，各骨格情報の両手首の座標が最も近づいたフレームを時刻の基準と定める．

### 3 研究結果

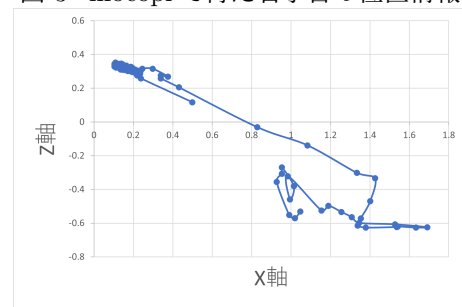
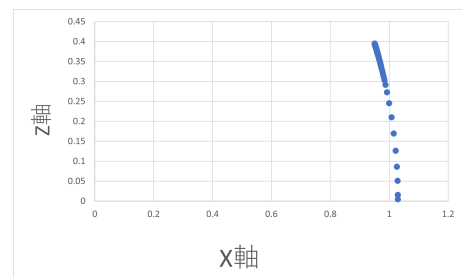
#### 3.1 実験方法

全身運動をしている一名の被験者(男子学生)が(2.3)に示す骨格情報のキャリブレーションに必要なポーズを行い，正しくキャリブレーションできているか骨格データを評価する．手を広げた状態から手を合わせるポーズになるまでは170フレームで行われている．手を広げたポーズを取ったとき，へそを原点，体の中心から右手首までの距離を1とした  $x$ - $z$  平面で表す．

#### 3.2 解析結果

mocopi により取得した右手首の骨格座標情報を図3，mediapipe により取得した右手首の骨格座標情報を図4に示す．

今回行った運動では，へそより前に右手首が存在しているため， $z$  軸方向の値が負になる娘はありえない．また，へその前まで腕を持ってきたので，正しく計測できていれば原点を中心とした半径位置の円に似た軌道を描くはずである．mocopi は奥行きに縮んでいることがわかる．ジェスチャーのような大まかな体の動きは分かるが，センサ位置の絶対的な座標は出していないことが分かった．RGB 画像による方法では，奥行きに歪んでいることに加え， $z$  軸方向の値が負になっている場面があるため， $z$  軸方向の計測値に大きな誤差があることが分かる．どちらも課題が残る結果になった．



### 4 まとめ

本研究では，キャリブレーションの方法として2つのポーズを続けて行い，3種類の手法を定量的に比較するためのキャリブレーションを試みた．RGB カメラによる方法では機械学習を用いて奥行き情報を取得できる方法を用いたが，慣性式モーションキャプチャデバイスで取得した骨格データより精度が下がることが確認できた．

RGBD による方法では現在問題の修正中であるが RGBD による方法を用いることで，デプス情報を元に正しく測定できるのではないかと考えられる．

### 参考文献

- [1] 剣 一輝，“柔道競技の 3D アーカイブ化”，令和 4 年度長岡高専専攻科論文，2023.
- [2] Google, “mediapipe”, <https://developers.google.com/mediapipe>
- [3] 北川リサ, 伊藤貴之, “競技かるたにおける払いの動作の三次元ボーン表示による可視化”，情報処理学会第 85 回全国大会，2023，1，139 – 140，2023.