

Abstract

全身運動を行う人物をカメラで撮影し、三次元骨格推定を行う二種類の方法を実装し比較する。一方は RGB 画像を入力として用いる方法で、他方は RGB 画像に加えデスプという、カメラと物体の奥行き情報を取得できるカメラを用いる方法である。また、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、定量的な評価を行った。RGB 画像による計測結果と慣性式モーションキャプチャデバイスによる計測結果を比べたとき、RGB 画像による計測結果では奥行きの推定結果に誤差が生まることがわかった。

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、アーカイブするために、カメラで撮影できる二次元の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^[1]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、近年カメラと物体までの奥行きの情報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^[3]。

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 RGBD を実装する。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には、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、カメラと物体の距離も取得可能な RGBD カメラを用いる方法がある (2.1)。

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す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 (2.2)。

本研究では、市販の入力デバイスを使用し、3つの方法について性能を比較評価する。

2.1 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(a) に示す 33 個の関節の三次元骨格情報取得できる。2つ目はカラー画像とデスプ情報を入力として 3DiVi Inc が提供するライブラリ NuiTrack を用いることで図 2(b) に示す 19 個の関節の三次元骨格情報が取得できる。

2.2 慣性式モーションキャプチャの三次元骨格推定

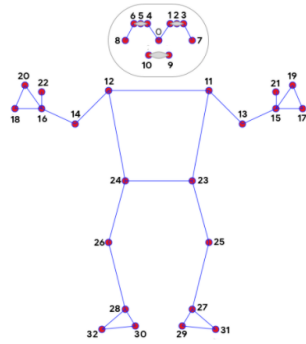
慣性式モーションキャプチャデバイスとして mocopi を用いる。mocopi のモーションデータのフレームレートは 60 fps である。両手、両足、頭、腰の計 6ヶ所に 3つの自由度を持つ角度センサと加速度センサを搭載した小型センサを装着してリアルタイムに三次元計測を行うことができる。6つの小型センサで計測しているため肘や膝などの関節部の屈曲を正確に表現することはできないが、機械学習を用いることで図 2(c) に示すような肘や膝などの関節部を含めた 27 個の関節位置を推定している。

2.3 キャリブレーション

まず、骨格情報によってスケールや基準となる座標系が違うので定量的に評価するためには統一する必要がある。全身運動を計測する前に、両腕を水平に上げるポーズをする。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



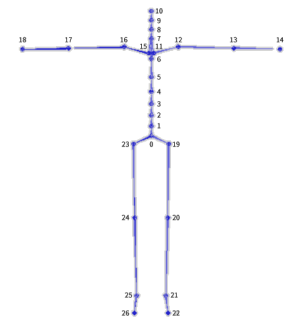
図 1 画像処理による骨格推定の様子



(a) MediaPipe Pose で取得できる関節位置



(b) NuiTrack で取得できる関節位置



(c) mocopi で取得できる関節位置

図 2 取得できる骨格

x 軸，これらに軸の直行する方向を z 軸とする。

次に時間遅れを修正する．両腕を水平に上げるポーズから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，両手首が最接近しているので，各骨格情報の座標が最も近づいたフレームを合わせる．

3 研究結果

3.1 実験方法

全身運動をしている人物一人に対して計測を行い，計測値を人体に当てはめるときに不自然でないか，肘や膝の関節部分が正しく計測されているかを比較する．計測をする前に，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を評価する．

3.2 解析結果

今回，RGB 画像により取得した骨格データと慣性式モーションキャプチャにより取得した骨格データのキャリブレーション時の右手の座標結果を xz 平面に並べた結果を図 3，4 に示す．

図 4 では，なめらかに右手の座標が動いているのがわかるのに対し，図 3 では右手の座標がなめらかではなく，手が合わさりオクルージョンが発生することで手首や腕が見えなくなる段階でデプス情報に当たる，z 軸方向の検出率が下がってしまうことがわかった．

4 まとめ

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，オクルージョンの発生に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，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”，

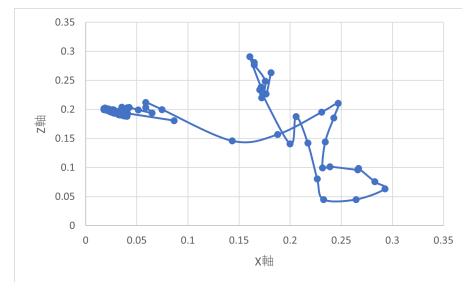


図 3 RGB 画像による方法で得た右手首の位置情報

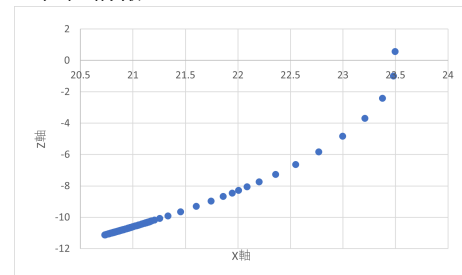


図 4 慣性式モーションキャプチャによる方法で得た右手首の位置情報

<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.

URL は 2023 年 10 月 4 日にアクセス