

## Abstract

a

## 1 研究背景・目的

情報通信技術や処理速度の向上により機械学習を活用して効率化を図る動きが様々な分野で広がっている。本研究室で行っている骨格推定の分野でも機械学習を活用し、少ない情報から三次元の骨格推定を行えるオープンソースが増えている。

そこで本研究では一台の入力装置で三次元骨格推定を行う機械学習を活用した現行の方法を行う。そして、動作の緩急やオクルージョンというカメラの手前にある物体に計測する箇所が隠れてしまい計測の信頼度が下がってしまう状態の有無について、それぞれの方法の精度を定量的に比較する。

## 2 研究内容

## 2.1 人の動作の計測方法

人の動作の三次元計測を行うには、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 カメラを用いて解析する方法 (2.2) と、カメラと物体の距離を測ることができる RGBD カメラを用いて解析する方法 (2.3) がある。

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推定する。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 RGB カメラ一台を用いた三次元骨格推定

一台の RGB カメラで撮影した動画から MediaPipe Pose<sup>[1]</sup> というライブラリを用いて処理することで骨格推定が行える。

MediaPipe は Google が提供しているライブメディアやストリーミングメディア向けの機械学習ソリューションである。顔認識や物体検出などが行えるライブラリがあり、計算コストが低くスマートフォンや組み込み PC など資源の限られたハードウェアでも使用できる。また、その中でも MediaPipe Pose は動画から人間の姿勢を推論するライブラリで、動画から図 1(a) に示す全身の 33 個の関節の三次元位置または上半身の 25 個の関節の三次元位置を予測できる。

## 2.3 RGBD カメラで行う三次元骨格推定

kinectSDK<sup>[2]</sup> のようなデバイス専用のソフトウェアや、Nitrack<sup>[3]</sup> などのオープンソースを開発ライブラリに用いて処理することによってアプリケーションを作成できる。RGBD カメラを用いて骨格推定をする際に一般的という理由から Nitrack を使用する。Nitrack は 3DiVi Inc が開発した三次元トラッキングミドルウェアで、RealSense D400 シリーズや Orbbec Astra シリーズ等の RGBD カメラで骨格推定が可能であり、図 1(b) に示す全身の 19 個の関節部分のトラッキングが可能である。

## 2.4 慣性式モーションキャプチャによる三次元骨格推定

画像処理による方法の精度を比較する際、画像処理とは独立した方法取得した骨格を基準としないと特定の方法に有利な結果が出てしまう可能性がある。そのため、画像処理とは独立した方法としてモーションキャプチャデバイス mocopi を用いて計測する。

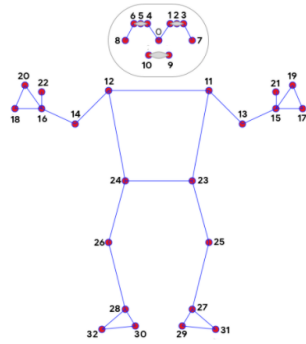
mocopi とは、市販のモーションキャプチャデバイスで両手、両足、頭、腰の計 6 か所に小型センサを装着してリアルタイムに三次元計測を行うことができる。6 つの小型センサで測定しているため肘や膝などの関節部の屈曲を正確に表現することはできないが、mocopi のセンサはそれぞれ 3 つの自由度を持つ加速度センサと角度センサで測定しており、機械学習を用いることで、1(c) に示すような肘や膝などの関節部を含めた全身の 27 個の関節位置を推定している。

## 2.5 キャリブレーション

計測方法により、座標系やスケール、処理による発生したフレームのズレが違う。そのため、座標系やスケール、処理による発生したフレームのズレを統一するためのキャリブレーションを行う必要がある。

キャリブレーションの方法として計測前に 2 つの動作を行う。1 つ目が両腕を水平に広げるポーズである。このポーズをしたときの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 z 軸、右手から左手に向かう方向を x 軸、これらの軸と直交する方向を y 軸として座標系を定める。

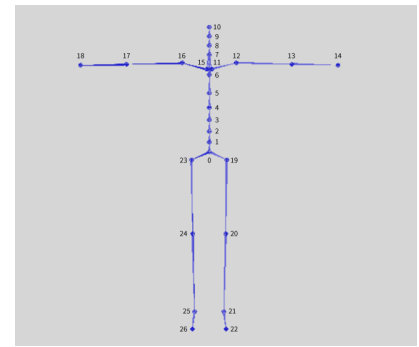
2 つ目が両手を胸の前であわせるポーズである。このポーズをしたときは両手首が最接近しているため、このタイミングを基準としてフレームのズレを合わせる。



(a) MediaPipe Pose で取得できる関節位置



(b) Nutrack で取得できる関節位置



(c) mocopi で取得できる関節位置

図 1 取得できる骨格

## 3 研究結果

### 3.1 実験方法

RGB, RGBD の情報を撮るカメラには Intel RealSense Depth Camera D415 を使用し、慣性式のモーションキャプチャには mocopi を使用した。カメラの解像度は  $640 \times 360$ px, フレームレートは 60fps である。実験には明るい室内で撮影した動画を使用した。

全身運動をしている人物一人に対して計測を行い、mocopi のセンサの位置に当たる両手、両足、頭、腰の計 6ヶ所に関して、画像処理を用いた三次元骨格推定で得られた座標との誤差を比較することで精度を評価する。

### 3.2 解析結果

## 4 まとめ

## 参考文献

- [1] Google, “mediapipe”,  
<https://developers.google.com/mediapipe>

URL は 2023 年 10 月 4 日にアクセス