

## Abstract

全身運動を行う人物を一台の RGB カメラや RGBD カメラで撮影し、それぞれのカメラに合った画像処理を用いた手法で三次元骨格推定を行う。そうして取得した三次元骨格推定の結果を精度や処理速度などについて比較し分析する。

## 1 研究背景・目的

情報通信技術の急速な進歩により人工現実感、拡張現実感、複合現実感などの応用が広がっている。感染症対策を契機にオンラインコミュニケーションも増加し、インターネット上の仮想共有空間であるメタバースが注目されている<sup>[1]</sup>。その理由の一つに、離れている相手にテキストや音声だけでなく身振りや動作などのノンバーバルな情報の伝達を行うことが容易であるという点がある。

三次元の仮想空間で自分の分身となるアバターを自由に操作するには、体の動きを計測する必要がある。画像処理による方法<sup>[2]</sup>や専用デバイスを装着する方法<sup>[3]</sup>などが試みられている。画像処理による方法で三次元の情報を取得するためには先行研究のような複数台のカメラを用いる方法<sup>[4]</sup>があるが、狭い室内であるなどの場所の制約や、限られた予算の中で実装したいという資金の制約などによって、複数台のカメラを用いる方法取るのが難しい場合がある。また、機械学習を用いることで一台のカメラや少ないセンサでも三次元骨格推定を可能にするオープンソースが増えて

いる。本研究では1台の入力装置で三次元骨格推定ができる現行の方法について調査及び実装を行い、動作の緩急やオクルージョンという自分の体に体の一部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。それぞれの方法について簡単にまとめたものを表1に示す。画像処理による方法では画像から人の骨格を推定することで人の動作を解析できる。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付け付けることで骨格を

表 1 動作を計測する方法の種類と特徴

	カメラ		モーションセンサ	
	RGB	RGBD	光学式	慣性式
センサ装着	不要	不要	必要	必要
外から撮影	必要	必要	必要	不要
必要台数	1～数台	1台	複数台	0台
計測方法	MediaPipe	Nuitrack		mocopi
計測関節数	33個	19個		27個

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骨格推定を行う場合、MediaPipe Pose<sup>[8]</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。特徴として、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という点が挙げられる。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

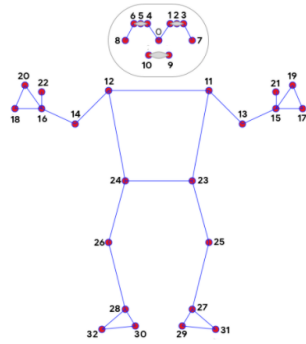
kinectSDK<sup>[6]</sup>のようなデバイス専用のソフトウェアや、Nuitrack<sup>[7]</sup>などのオープンソースを開発ライブラリに用いて処理することによってアプリケーションを作成できる。

代表的な RGBD カメラと開発ライブラリの組み合わせを表 1b に示す。

本研究で扱う Nuitrack 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、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

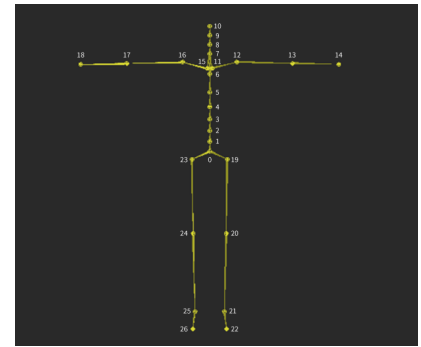
画像処理の精度を比較する際、画像処理による方法で取得したデータを基準にするのは特定の骨格推定の方法に有利な結果が出てしまう可能性がある。そのた



(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

め、基準とするデータは画像処理に頼らない独立した方法で取得する必要があるなどの理由からモーションキャプチャデバイス mocopi<sup>[9]</sup> を用いる。

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、肘や膝などの関節部を含めた全身の推定を実現している。

## 2.5 キャリブレーション

計測方法によりスケール、処理により発生する遅延によるフレームのズレ、座標系が異なるため統一しなければ定量的な比較が行えない。そのため、これら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、計測する運動をする前に両手を水平に広げその時の両手首の長さを元にスケールを合わせる。次に、両手を合わせ両手首の座標が最接近した時点を基準としてフレームのズレを合わせる。最後に、へその座標を原点として頭に向かう方向を z 軸、右手首から左手首に向かう方向を x 軸、これらの軸と直行する方向を y 軸として座標系を定める。

## 3 研究結果

### 3.1 実験方法

mocopi を装着し、

表 2 右手首の基準と測定値の差分

オクルージョン	動作が緩やか 動作が急			
	なし	あり	なし	あり
RGB				
RGBD				

## 3.2 実験結果

## 4 まとめ

## 参考文献

- [1] 日本経済新聞, “孤独感抱える若者の交流の場、メタバースに開設”, <https://onl.la/Y7BehPP>
- [2] 平尾 公男ら, “多関節 CG モデルと距離画像による上半身の姿勢推定”, Technical report of IEICE. PRMU, VOL.104, No.573, 79–84, 2004.
- [3] 白鳥 貴亮ら, “モーションキャプチャと音楽情報を用いた舞踊動作解析手法”, 電子情報通信学会論文誌 D, Vol.J88-D2, No.8, pp.1662-1671, 2005.
- [4] 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.
- [5] 安達 康平ら, “ビデオからの 3 次元姿勢を用いた行動認識における精度向上の試み”, 研究報告モバイルコンピュータビジョンとパーベイシブシステム (MBL), 2020-MBL-94, 47, 1–7, 2020.
- [6] 谷尻 豊寿, “体の動きがコントローラ C++で kinect プログラミング KINECT センサ画像処理プログラミング”, 株式会社 カットシステム, 2011.
- [7] 3DiVi Inc, “Nuitrack”, <https://nuitrack.com/jp>
- [8] Google, “mediapipe”, <https://developers.google.com/mediapipe>
- [9] SONY, “モバイルモーションキャプチャー mocopi”, <https://www.sony.jp/mocopi/>

URL は 2023 年 10 月 4 日にアクセス