

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。また、それ以外にも一台のカメラと画像処理を用いて三次元骨格推定を行う方法がないかリサーチする。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^[1]。

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^[2]や専用デバイスを装着する方法^[3]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^[4]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。

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、それぞれの方法の計測速度や精度について比較する。

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などがある。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を取り付けそのマーカ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。

本研究では一台のカメラと画像処理で三次元計測を行う複数の方法について検証し、モーションセンサを用いて三次元計測した場合と比較する。

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

2.2 RGB カメラ一台を用いた三次元骨格推定

RGB カメラで撮影した動画を MediaPipe Pose^[7]というライブラリを用いて処理することで骨格推定を行う。

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、少ない記述量で使うことができ、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ものが挙げられる。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、動画から図1に示す全身の33個のランドマーク位置または上半身の25個のランドマーク位置を予測できる。

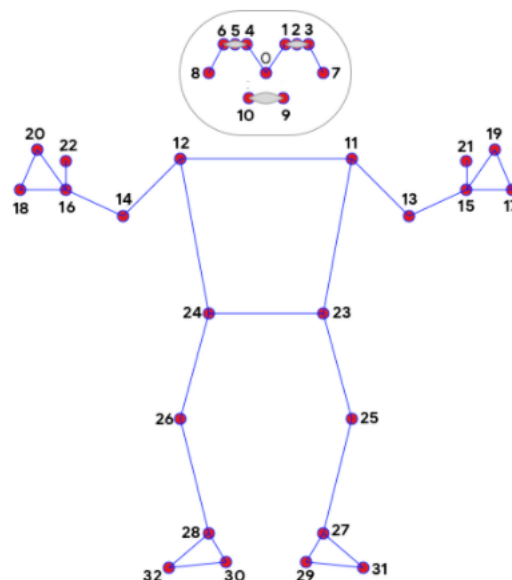


図1 MediaPipe Pose で取得できるランドマーク

表 2 RGBD カメラと開発ライブラリの組み合わせ

カメラ	開発ライブラリ
kinect2	kinectSDK, OpenNI
RealSense	Nuitrack

2.3 RGBD カメラで行う三次元骨格推定

入力を RGBD カメラにする場合、kinectSDK^[6]などのハードウェアの公式から出ているソフトウェアや mediapipe^[7]などのオープンソースを処理に用いることによって三次元骨格推定を行うことができる。RGBD カメラを入力装置として用いる際の関係を図 2 に示す。kinectSDK^[6]のようなデバイス専用のソフトウェアや、Nuitrack^[7]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 カメラと開発ライブラリの組み合わせを表 2 に示す。

kinectSDK は Microsoft Kinect for Windows Software Development Kit の略で、Microsoft 社から公式にリリースされた開発キットである。Windows 上で kinect2 を動かすのに必要なドライバやドキュメントなどが同梱されている。kinectSDK はハードウェアである kinect2 との組み合わせに限定されており、現在は開発が終了してしまっている。

kinectSDK と似たような機能を持つもので、Kinect のセンサ部分を開発した PrimeSense 社が中心となって開発した OpenNI というライブラリがある。OpenNI には部分的なトラッキングやジェスチャ検出機能など kinectSDK に比べることが多く、GitHub 上でソースコードが公開されており現在も有志による開発が行われている。

Nuitrack は RealSense D400 シリーズや Orbbec Astra シリーズ等の RGBD カメラで骨格認識を可能にするミドルウェアで、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 画像処理データの比較方法

各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ため、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^[8]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。

本研究では、ラジオ体操のような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 か所に関して、画像処理を用いて行った三次元骨格推定で得られた座標との誤差や時間変動を比較することで精度を評価する。また、計測する運動は動作の緩急とオクルージョンというカメラに対し手前にある物体が後ろにある物体を隠す状態になり十分に計測できない場合の有無についてそれぞれの動作を満たす物を抜粋する。

3 研究計画と進捗状況

mocopi を装着してラジオ体操をしている人を RealSense で撮影し、そのデータを用いて骨格推定を行う。推定した骨格と mocopi で測定した骨格の両手、両足、頭、腰の座標のズレを誤差として精度の計測を行う。

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ティングとパーベイスシステム (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] Google, “mediapipe”, <https://developers.google.com/mediapipe>
- [8] 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

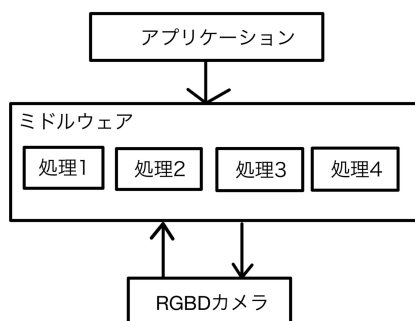


図 2 OpenNI の構成