

Abstract

本文が完成してから書きます。

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。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や専用デバイスを装着する方法などが試みられている。特に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複数台のカメラを用いる方法があるが、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予算の制約などによってこの方法を取るのが難しい場合がある。

本研究ではカメラ1台で三次元骨格推定ができる現行の方法について比較し、それぞれの方法のメリットやデメリット、精度などについて比較する。また、それらを元に組み込みPCでの実装やリアルタイム処理などの高速化、オクルージョンへの対応などの高精度化を目指す。

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とモーションセンサによる方法がある。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知ることができる。画像処理によって三次元骨格推定するには、色の情報であるRGBの情報を取れるごく一般的なRGBカメラを用いて撮影する方法と、kinectやRealSenseのようなカメラと物体の距離を測ることができるRGBDカメラを用いて撮影する方法がある。

今回は場所や予算などの制約から一台のカメラで人の動作の三次元計測を行う場合に限定して比較するので、先行研究^[1]のような複数台のカメラを用いて三次元計測を行う場合を除いた方法について検証する。

2.2 一台のカメラで行う三次元骨格推定

本研究では、RGBカメラを用いる方法、kinect2を用いる方法、intel RealSenseを用いる方法の3つの方法で三次元骨格推定を行う。

RGBカメラを用いる方法ではOpenPoseと3d-pose-baselineを用いて、以下の手順で三次元骨格推定を行うことができる。

1. 2000万画素、30fpsのGoProで撮影
2. 撮影した動画を連続静止画へ変換
3. 各静止画からOpenPoseで関節の二次元位置を抽出
4. 関節の二次元位置を時間軸方向に平滑化
5. 関節の二次元位置を3d-pose-baselineの入力形式に変換
6. 3d-pose-baselineで三次元位置を推定

OpenPose^[5]とは、カーネギーメロン大学のCaoらによって発表された、18個のキーポイント（関節）とその関節をつなぐボーン（骨）を検出することができるオープンソースである。OpenPoseは、正面からの画像だけでなく横からでも姿勢推定を行うことができる。また、信頼度は低下するが遮蔽物により、見えない部位の推定も行うことができる。

3d-pose-baseline^[2]は三次元の姿勢情報と二次元に投影した姿勢情報を機械学習することによって、二次元の姿勢推定情報から三次元骨格推定が行え、その座標情報を取得できるものである。

kinect2を用いる方法ではkinectSDKを使うことでkinect2に搭載されているセンサーを制御してRGB画像や奥行き情報をPCに入力することができる。また、それらの情報から人体の骨格を自動的に検出し、その骨格を追跡することができる。

kinectSDK^[3]とは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、Windows上でkinect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。似たような機能を持つもので、Kinectのセンサー部分を開発した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。こちらにはセンサー部分でできることが多いが、ドライバが同梱されておらず一つのライブラリで完結しないので今回はKinectSDKを使用することとする。

RealSenseを用いる方法では、cubemosのSkeleton Tracking SDKを使うことによって三次元骨格推定をすることができる。cubemosのSkeleton Tracking SDKはディープラーニングベースの2D/3D骨格トラッキング機能を、組み込みハードウェア向けのアプリケーションへ提供するように設計されたソフトウェアである。そのため、GPUがなくても動き、WindowsやLinux上で動かす、対応言語が比較的多いという特徴がある。

2.3 精度比較

三種類の現行の方法を比較する際、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それぞれ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行う必要がある。そこで、加速度センサと角度センサが搭載されているモーションキャプチャデバイス mocopi を用いて取得した骨格データとの誤差を元に精度を比較する。

mocopi^[4]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。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、AIを利用して人の様々な動作を予め学習させておくことで、センサを装着していない肘や膝などの中間関節を含めた全身の推定を実現している。よってmocopiでは肘や膝などの関節の屈曲を正確に表現することはできない。そこで本研究では、オクルージョンが起きにくく動きが早すぎない動作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を比較することで精度を評価する。

3 研究計画と進捗状況

3.1 研究の進め方

mocopiを装着して学校体操をしている人をRGBカメラ,kinect2,RealSenseで撮影し、それぞれのカメラで撮影した映像から三次元骨格推定を行い、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。

3.2 研究方法や装置の概略

本研究では撮影機材としてRGBカメラ, kinect2, RealSense, mocopiを用いる。また、開発環境としてOpenPose, 3d-pose-baseline, kinectSDK, cubemos Skeleton Tracking SDKを使用する。

3.3 進捗状況

現在は、OpenPoseによる姿勢推定を進めている。今後は記述した方法だけでなく、他にも単一のカメラで三次元骨格推定ができる方法がないかリサーチしつつ、高精度化、高速化、組み込みPCでの実装を目指していく。

参考文献

- [1] 剣 一輝, “柔道競技の3Dアーカイブ化”, 令和4年度専攻科修士論文, 2023年
- [2] J. Martinez, R. Hossain, J. Romero, J. Little. “A simple yet effective baseline for 3d human pose estimation”. In ICCV, 2017年
- [3] 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 センサー画像処理プログラミング”, 株式会社 カットシステム, 2011年
- [4] モ バ イ ル モ ー シ ョ ン キャ プ チ ャ ー mocopi, <https://www.sony.jp/mocopi/>
- [5] CAO,Zhe,et al.OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. arXiv preprint arXiv:1812.08008. 2018.