B – 画像処理による人物の全身動作の三次元骨格推定方法の比較

35 番　本間 三暉（視覚情報処理研究室/高橋 章）

Abstract

運動を行う人物を一台の RGB カメラや RGBD カメラで動画を撮影し，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．それらの方法で行った三次元骨格推定を比較し，精度や処理速度の観点から比較分析する．

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．メタバースが注目されている理由の一つに，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り，画像処理による方法 [1] や専用デバイスを装着する方法 [2] 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 [3] 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実験を行い，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，それらの情報を元に組み込み 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への対応を目指す．

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することができる．画像処理によって三次元骨格推定するには撮影するカメラに，色情報を記録する一般的な RGB カメラで撮影して解析する方法と，カメラと物体の距離を測ることができる RGBD カメラで撮影し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等があるが，どの方法もマーカーやセンサを検出対象に取り付けなければならないため使用できる環境が限定されてしまう．

本研究では一台のカメラと画像処理によりで三次元計測を行う場合について検証する．

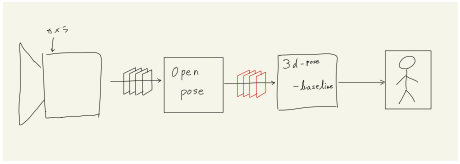

図 1　単一 RGB カメラによる三次元骨格推定の流れ

2.2　一台の RGB カメラで行う三次元骨格推定

RGB カメラで撮影して解析する方法では処理に 3d-pose-baseline [4] と OpenPose [9] を用いる．

OpenPose 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 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．

以下の手順で三次元骨格推定を行う [5] ことができる．簡単に図にまとめたものを図 1 に示す．

一般的な RGB カメラで運動している人物の動作を撮影し，その動画を連続静止画へ変換する．各静止画から OpenPose で関節の二次元位置を抽出する．

抽出した二次元位置を三点移動平均など用いて滑らかな動作をしているように整える．そうして動きをなめらかにした関節の二次元位置を 3d-pose-baseline の入力形式に変換し，3d-pose-baseline を用いて三次元位置を推定する．

2.3　RGBD カメラで行う三次元骨格推定

入力を kinect2 や intel の RealSense などの RGBD カメラにする場合，処理に kinectSDK [6] などの公式から出ているや mediapipe [7] などのオープンソースを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．

kinectSDK 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て，ソフトウェアである kinectSDK とハードウェアである kinect があれば最低限動く様になっ

表 1　動作を計測する方法の種類と特徴

動作を計測する方法の特徴についてまとめたものの表

ている．

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．こちら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 に比べできることが多く，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が，OpenNI 自体はミドルウェアの位置付けでありソフトウェアとハードウェアの架け橋に過ぎないのでこれだけで動くことはない．だが，kinectSDK は現在開発が終了してしまっているので，どちらにも長所と短所が存在する．

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，少ない記述量で使う事ができて，バッテリー駆動のデバイスでも十分に動作するよう設計されているというものが挙げられる．

kinect2 で撮影する場合は kinektSDK や OpenNI を用いて解析することで三次元骨格推定を行うことができる．

RealSense で撮影する際，medeiapipe [7] の Pose 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．

2.4　比較方法

各画像処理の精度を比較する際，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折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，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．

mocopi を用いる理由として，現行の RGB カメラや RGBD カメラで撮影し処理を行うような方法を比較する際，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り，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ことや，人体の左右対称性を考慮した骨格データが出てくるためなどが挙げられる．

そこで本研究では，学校体操やラジオ体操のような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や時間変動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

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．

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，オクルージョンへの対応，高速化，組み込み PC での実装を目指していく．

参考文献

[1] 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.

[2] 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.

[3] 剱 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.

[4] J. Martinez, R. Hossain, J. Romero, J. Little. “A simple yet effective baseline for 3d human pose estimation” . In ICCV, 2017.

[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] 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/

[9] CAO,Zhe,et al.OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. arXiv preprint arXiv:1812.08008. 2018.