令和 5 年度電子制御工学科卒業論文

単一視点による
人物の全身運動の三次元計測について

長岡工業高等専門学校
電子制御工学科
視覚情報処理研究室
(指導教員 高橋　章教授)

本間　三暉

2024/02/22

目次

第 1 章

はじめに

人の動きなどのノンバーバルな情報をコミュニケーションに用いたり，技術の継承や伝統芸能のディジタルアーカイブに利用したりするために，カメラの二次元画像から三次元の骨格情報を推測する技術が求められている．本研究室では柔道の三次元の動きを複数視点の動画から推測する研究 [1] が行われた．しかし，この方法では複数台のカメラを設置し，キャリブレーションを行うため，十分な広さを持つ測定空間が必要である．また，近年技術開発が進んだことにより，物体までの奥行きの情報（デプス）を取得できるカメラが市販されたり，機械学習によって単眼カメラから深度推定を行う方法 [2] が提案されたりしている．

そこで本研究では機械学習を活用して一台の入力装置で三次元骨格推定を行う手法として，RGB 画像を入力する方法と RGB 画像に加えてデプス情報を入力する方法を実装する．さらに慣性式モーションキャプチャデバイスによる計測結果と比較するための座標系及び時間のキャリブレーション方法を検討する．

第 2 章

研究内容・方法

人の動作の三次元骨格推定を行う方法として，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 2.1 に示す．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 カメラを用いる方法と，物体のデプスも取得可能な RGBD カメラを用いる方法がある (2.2 節)．

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高精度に推定できるが広い計測空間が必要になる．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 (2.1 節)．

本研究では，市販の入力デバイスを使用する 3 つの推定方法を実装して性能を比較評価する．

表 2.1: 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	0

2.1 慣性式モーションキャプチャの三次元骨格推定

慣性式モーションキャプチャデバイスとして mocopi[3] を用いる．mocopi のセンサは 3 つの自由度を持つ角度センサと加速度センサを搭載している．両手，両足，頭，腰の計 6 ヶ所に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．センサを装着しない肘や膝などの関節部を直接測定することはできないが，機械学習を用いることで p.5 の図 2.1(a) に示すような 27 個の関節位置を推定している．

mocopi は専用のスマホアプリでモーションを収録することで BVH ファイル (付録 A を参照) での書き出しが可能である．この時 mocopi のモーションデータのフレームレートは 30，50，60 fps から選ぶことが可能である．書き出した BVH ファイルはホストとなるスマホの機種により保存されるフォルダが異なり，iphone の場合 mocopi フォルダの MotionData 配下，Android スマホの場合ユーザーが初回保存時に指定したフォルダ配下に保存される．

本研究ではホストとなるスマホに iPhone SE（第 2 世代）と Google Pixel 8 を用いた．しかし，Google Pixel 8 ではモーションセンサとの接続が不安定となってしまうため，測定には iPhone SE（第 2 世代）をホストとした．また，モーションデータのフレームレートは 60 fps とした．

2.2 画像処理による三次元骨格推定

画像処理を用いて三次元骨格推定をしている様子を p.6 の図 2.3 に示す．本研究では画像入力装置として RGBD カメラである Intel RealSense D415（以下 RealSense）を使用する．RealSense のスペックを付録 B に示す．

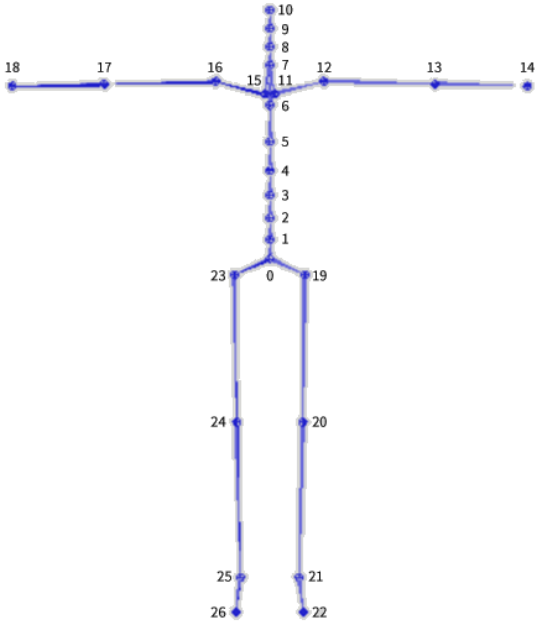
RealSense はカメラから物体までの距離を図る方法として，RealSense 内部のプロジェクタとカメラを用いてアクティブステレオ法を行っている．アクティブステレオ法とは，プロジェクタから投影するパターンの画像と，カメラの画像との間でマッチングを行うことで距離を測定する方式である．投影するパターンは不可視の赤外線を利用している．

RealSense から手に入れた RGBD 画像から RGB 画像を取り出し, RGB 画像と RGBD 画像それぞれを入力とする．2 種類の入力画像からそれぞれ画像処理を用いて骨格推定をする方法として 2 つの方法を実装する．

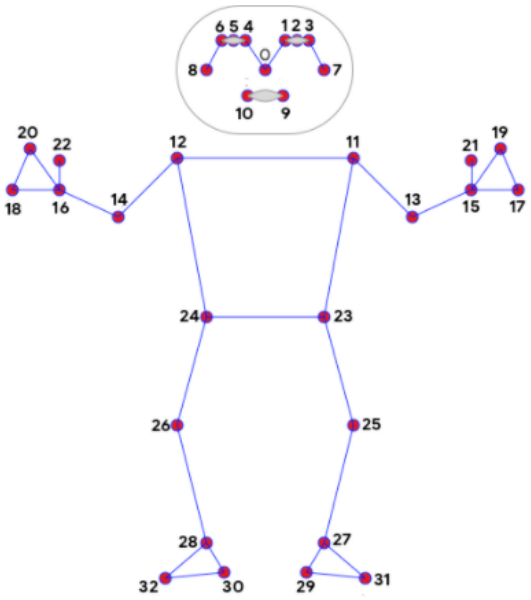
1 つ目はカラー画像を入力として Google が提供するオープンソースの機械学習ライブラリ MediaPipe Pose[6] を用いることで図 2.1(b) に示す 33 個の関節の三次元骨格情報

を取得する方法である．MediaPipe Pose は三次元の骨格情報と二次元画像を関連付けたデータを元に機械学習をすることで，二次元画像から三次元の関節の位置情報を推定することを可能にしている．実装した様子を p.6 の図 2.2(b) に示す．

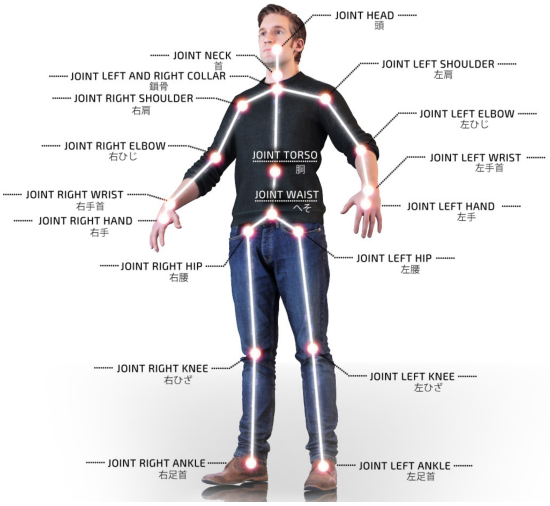
2 つ目はカラー画像とデプス情報を入力として 3DiVi Inc が提供するライブラリ Nuitrack[7] を用いることで図 2.1(c) に示す 19 個の関節の三次元骨格情報を取得する方法である．Nuitrack は機械学習と 3DiVi 社独自のアルゴリズムによって，RGBD 画像から三次元の関節の位置情報を推定できる．実装した様子を p.6 の図 2.2(c) に示す．



(a) mocopi で取得できる関節位置

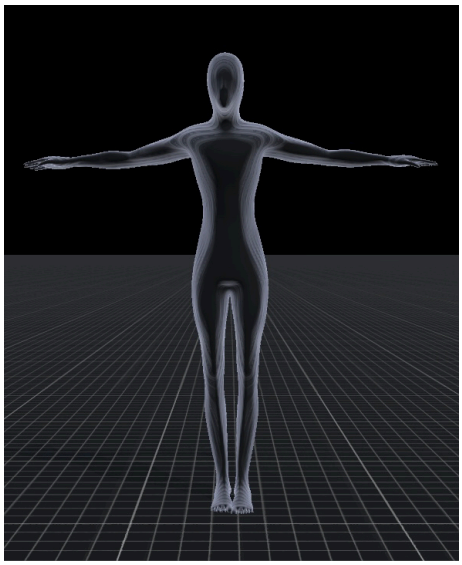


(b) MediaPipe Pose で取得できる関節位置



(c) Nuitrack で取得できる関節位置

図 2.1: 各ライブラリで取得できる骨格の関節位置



(a) mocopi で取得できる関節位置

(b) MediaPipe Pose で骨格推定した様子

(c) Nuitrack で骨格推定した様子

図 2.2: 各ライブラリで骨格推定した様子

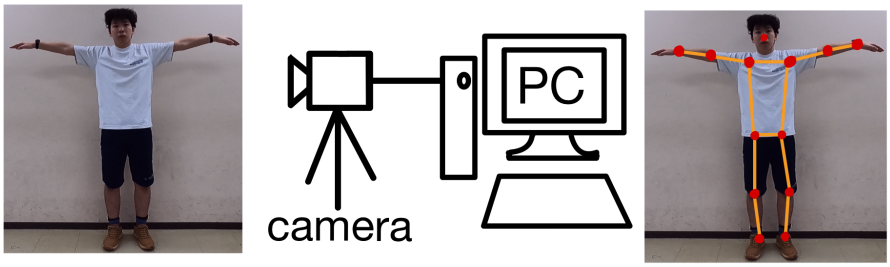


図 2.3: 画像処理による骨格推定の様子

2.3 キャリブレーション

推定方法によって骨格座標のスケールや基準となる座標系が違うので，定量的に比較評価するためには座標系を統一する必要がある．そこで計測開始時に，両腕を水平に上げるポーズを取ることにする．水平に上げた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 y 軸，右手から左手に向かう方向を x 軸，これらに軸の直行する方向を z 軸と定める．

動作の比較を行うには同期を取る必要があるが，画像処理による方法と mocopi を用いる方法では互いの同期計測ができない．そこで座標系を合わせるポーズの後で両手を伸ばしたまま胸の前で合わせるポーズをする．手が合わさっている時，両手首が最接近するので，各骨格情報の両手首の座標が最も近づいたフレームを時刻の基準と定める．

第 3 章

研究結果

3.1　実験方法

一名の被験者 (男子学生) が 2.3 節に示す骨格情報のキャリブレーションに必要なポーズを行い，正しくキャリブレーションできているか評価する．座標軸のキャリブレーションのポーズである手を広げた状態から，時間軸の基準となる両手が最接近した状態になるまでは 170 フレームで行われている．手を広げたポーズを取ったとき，へそを原点，体の中心から右手首までの距離を 1 とした x-z 平面で表す．

3.2　解析結果

Nuitrack による方法では関節の位置座標を取得することが困難であったため，mocopi と MediaPipe Pose による方法で取得した骨格座標情報を示す．mocopi により取得した右手首の骨格座標情報を図 3.1，MediaPipe Pose により取得した両手首の骨格座標情報を図 3.2 に示す．

今回行ったキャリブレーションのポーズでは，カメラから見てへそより前に両手首があるので，z 軸方向の値が負になることはありえない．また，へその前まで腕を持ってきているので，正しく計測できていれば両手首の座標を x-z 平面に示した際の軌跡は原点を中心とした半径 1 の円に似た軌道を描くはずである．

mocopi は実際の骨格の三次元位置より，動作の滑らかさを優先していることがわかった．動かした際の人間らしくない動きをなくすためではないかと考えられる．

MediaPipe Pose は多少のノイズがあるが，体に対し平行な向きはある程度正確に測定できることがわかった．しかし，奥行き方向に歪んでいることもわかった．

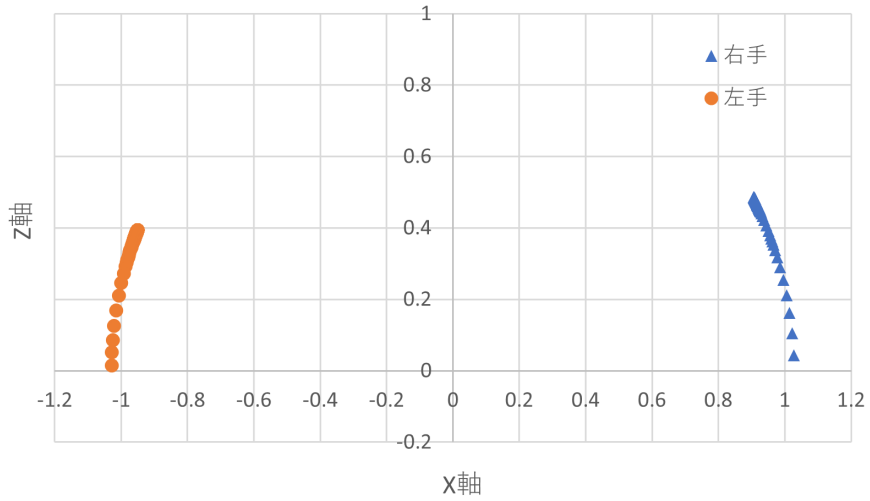


図 3.1: mocopi で得た両手首の位置情報

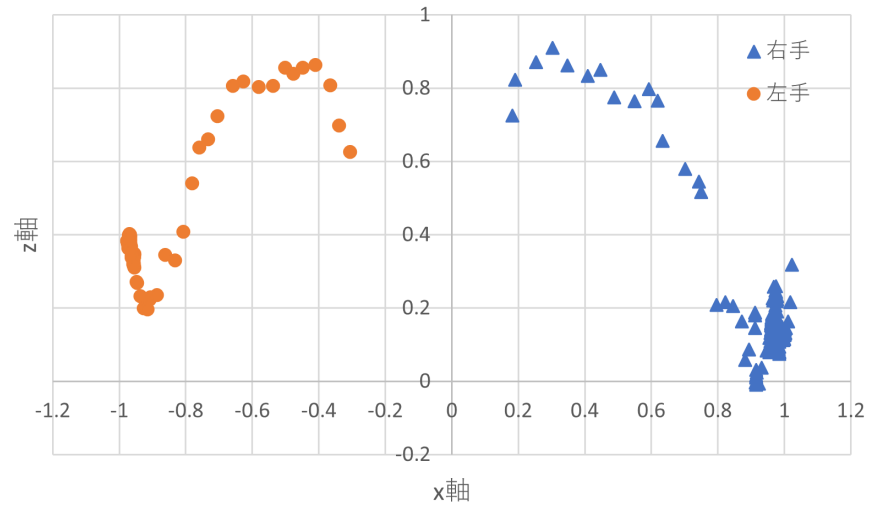


図 3.2: MediaPepe Pose で得た両手首の位置情報

第 4 章

まとめ

本研究では，キャリブレーションの方法として 2 つのポーズを続けて行い，3 種類の手法を定量的に比較するためのキャリブレーションを試みた．また，mocopi による方法では各骨格の三次元位置より動作の滑らかさを優先している事がわかった．RGB カメラによる方法では機械学習を用いて奥行き情報を取得できる方法を用いたが，慣性式モーションキャプチャデバイスで取得した骨格データよりノイズが大きくなることが確認できた．

今後の課題として，Nuitrack を用いて三次元骨格推定を行った場合の関節の位置座標を取得しキャリブレーションを行う．また，今回実装した 3 種類の方法を比較した際に，それぞれどのような特徴がどのような理由で出るのか考察を行う．mocopi の骨格データを補正する事はできないか検証を行うことなどが挙げられる．

謝辞

本研究におきまして，視覚情報処理研究室の高橋章教授から多大なるご指導，ご助力を賜りましたことを深く感謝するとともに，心から御礼申し上げます．また，ご協力いただいた同研究室の各氏にも感謝します．

参考文献

[1] 劔 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.

[2] 北川リサ，伊藤貴之，“競技かるたにおける払いの動作の三次元ボーン表示による可視化”，情報処理学会第 85 回全国大会，2023，1，139 – 140，2023.

[3] SONY，“モバイルモーションキャプチャー mocopi”，https://www.sony.jp/mocopi/

[4] TMPSwiki，“MOCAP データファイル”,https://mukai-lab.org/content/MotionCaptureDataFile.pdf

[5] Intel, “Depth Camera D415”, https://www.intelrealsense.com/depth-camera-d415/

[6] Google，“MediaPipe”，https://developers.google.com/mediapipe

[7] 3DiVi，“Nuitrack SDK”，https://www.aerotap.com/nuiTrack/doc/Overview_page.html

[8] L.D. ランダウ，E.M. リフシッツ，“力学”，pp.138–139

付録 A

mocopi から出力される BVH ファイルの構造

BVH ファイルはモーションキャプチャデータファイルフォーマットである．BVH ファイルの特徴を以下にまとめる [4].

- テキスト形式で記述
- 右手座標系で，xyz 各軸の扱い（どの軸が鉛直方向に対応するか等）はファイルの仕様としての指定はない
- 関節ノードに関する情報を記述
- 関節回転はオイラー角形式 (付録 C を参照) で記述
- 位置の単位は cm，回転角度の単位は degree
- スケルトン階層を表す HIERARCHY と，動作データを表す MOTION の二つから構成

mocopi から出力された BVH ファイルは，右手座標系で Y 軸が座標系内の鉛直方向としている．mocopi で計測したモーションデータの出力ファイルである BVH ファイルをリスト A.1 に示す.

まず，HIERARCHY のキーワードで始まる部分でスケルトンの構造を定義している．各関節のボーンの長さと初期方向，関節の親子関係，関節自由度の情報が記述されている．

ROOT は階層構造の始点となり，OFFSET，CHANNELS を要素に持ち，必ず 1 つ以上の JOINT もしくは End を持つ．JOINT は OFFSET，CHANNELS を要素に持ち，必ず 1 つ以上の JOINT もしくは End を持つ．End は階層構造の末尾となる特殊な JOINT である．OFFSET は親ノードから子ノードへの三次元の相対的な初期位置で，

ROOT の場合座標系内の絶対的な初期位置である．CHANNEL は，関節自由度に続き，位置の自由度 (Xposition, Yposition, Zposition)，回転の自由度 (Xrotation, Yrotation, Zrotation) を，ROOT から子ノードに向かう順序に定義する．

次に MOTION のキーワードで始まる部分で総フレーム数，1 フレームあたりの時間，モーションデータの順で記述する．モーションデータは各フレームごとに一行ずつ，各行にはファイルの冒頭からの CHANNEL の登場順に対応する．

リスト A.1: mocopi の BVH ファイル

```
1  HIERARCHY
2  ROOT root
3  {
4    OFFSET 0 90.5966 0
5    CHANNELS 6 Xposition Yposition Zposition Zrotation Xrotation
     Yrotation
6    JOINT torso_1
7    {
8      OFFSET 0 4.93189 -1.1181
9      CHANNELS 6 Xposition Yposition Zposition Zrotation Xrotation
     Yrotation
10     JOINT torso_2
11     {
12       OFFSET 0 5.45428 1.04046
13       CHANNELS 6 Xposition Yposition Zposition Zrotation Xrotation
     Yrotation
14         JOINT torso_3
15         {
16           OFFSET -9.96389e-18 5.81666 0.111686
17           CHANNELS 6 Xposition Yposition Zposition Zrotation Xrotation
      Yrotation
18           JOINT torso_4
19           {
20             OFFSET -1.07264e-17 6.3178 -0.430704
21             CHANNELS 6 Xposition Yposition Zposition Zrotation
     Xrotation Yrotation
22             JOINT torso_5
23             {
24               OFFSET -1.91593e-17 7.38567 -1.4624
```

```
25              CHANNELS 6 Xposition Yposition Zposition Zrotation
    Xrotation Yrotation
26              JOINT torso_6
27              {
28                OFFSET -2.45986e-17 9.22288 -0.888062
29                CHANNELS 6 Xposition Yposition Zposition Zrotation
    Xrotation Yrotation
30                JOINT torso_7
31                {
32                  OFFSET -2.82008e-17 10.2464 1.53133
33                  CHANNELS 6 Xposition Yposition Zposition Zrotation
    Xrotation Yrotation
34                  JOINT neck_1
35                  {
36                    OFFSET -1.25459e-17 4.64811 0.694664
37                    CHANNELS 6 Xposition Yposition Zposition
    Zrotation Xrotation Yrotation
38                    JOINT neck_2
39                    {
40                      OFFSET -1.75924e-17 4.68175 0.4096
41                      CHANNELS 6 Xposition Yposition Zposition
    Zrotation Xrotation Yrotation
42                      JOINT head
43                      {
44                        OFFSET -1.9515e-17 4.6908 0.74295
45                        CHANNELS 6 Xposition Yposition Zposition
    Zrotation Xrotation Yrotation
46                        End Site
47                        {
48                          OFFSET 0 0.1 0
49                        }
50                      }
51                    }
52                  }
53                  JOINT l_shoulder
54                  {
55                    OFFSET 1.19736 -7.38744 7.31601
56                    CHANNELS 6 Xposition Yposition Zposition
```

```
Zrotation Xrotation Yrotation
                    JOINT l_up_arm
                    {
                      OFFSET 12.5803 3.15559 -3.17425
                      CHANNELS 6 Xposition Yposition Zposition
Zrotation Xrotation Yrotation
                      JOINT l_low_arm
                      {
                        OFFSET 28.3639 0.0580795 0.133078
                        CHANNELS 6 Xposition Yposition Zposition
Zrotation Xrotation Yrotation
                        JOINT l_hand
                        {
                          OFFSET 23.5335 0.0481878 0.110415
                          CHANNELS 6 Xposition Yposition Zposition
Zrotation Xrotation Yrotation
                          End Site
                          {
                            OFFSET 0.1 0 0
                          }
                        }
                      }
                    }
                    JOINT r_shoulder
                    {
                      OFFSET -1.19736 -7.38743 7.31591
                      CHANNELS 6 Xposition Yposition Zposition
Zrotation Xrotation Yrotation
                      JOINT r_up_arm
                      {
                        OFFSET -12.5803 3.15559 -3.17425
                        CHANNELS 6 Xposition Yposition Zposition
Zrotation Xrotation Yrotation
                        JOINT r_low_arm
                        {
                          OFFSET -28.3639 0.0580795 0.133078
                          CHANNELS 6 Xposition Yposition Zposition
```

```
Zrotation Xrotation Yrotation
                        JOINT r_hand
                        {
                          OFFSET -23.5335 0.0481876 0.110415
                          CHANNELS 6 Xposition Yposition Zposition
Zrotation Xrotation Yrotation
                          End Site
                          {
                            OFFSET -0.1 0 0
                          }
                        }
                      }
                    }
                  }
                }
              }
            }
          }
        }
      }
    }
    JOINT l_up_leg
    {
      OFFSET 8.96978 -4.08401 1.97395
      CHANNELS 6 Xposition Yposition Zposition Zrotation Xrotation
Yrotation
      JOINT l_low_leg
      {
        OFFSET -0.827718 -37.6515 -0.63625
        CHANNELS 6 Xposition Yposition Zposition Zrotation Xrotation
Yrotation
        JOINT l_foot
        {
          OFFSET -2.25899 -39.2341 -5.87393
          CHANNELS 6 Xposition Yposition Zposition Zrotation Xrotation
 Yrotation
          JOINT l_toes
          {
```

```
122             OFFSET 0.756579 -9.59872 12.1348
123             CHANNELS 6 Xposition Yposition Zposition Zrotation
    Xrotation Yrotation
124             End Site
125             {
126               OFFSET 0 0 0.1
127             }
128           }
129         }
130       }
131     }
132     JOINT r_up_leg
133     {
134       OFFSET -8.96978 -4.08401 1.97395
135       CHANNELS 6 Xposition Yposition Zposition Zrotation Xrotation
    Yrotation
136       JOINT r_low_leg
137       {
138         OFFSET 0.827718 -37.6515 -0.63625
139         CHANNELS 6 Xposition Yposition Zposition Zrotation Xrotation
    Yrotation
140         JOINT r_foot
141         {
142           OFFSET 2.25899 -39.2341 -5.87393
143           CHANNELS 6 Xposition Yposition Zposition Zrotation Xrotation
     Yrotation
144           JOINT r_toes
145           {
146             OFFSET -0.756579 -9.59872 12.1348
147             CHANNELS 6 Xposition Yposition Zposition Zrotation
    Xrotation Yrotation
148             End Site
149             {
150               OFFSET 0 0 0.1
151             }
152           }
153         }
154       }
```

```
155     }
156   }
157   MOTION
158   Frames: 1193
159   Frame Time: 0.016667
160   0 90.5966 0 -0 0 -0 0 4.93189 -1.1181 -0 0 -0 0 5.45428 1.04046
        -0 0 -0 -9.96389e-18 5.81666 0.111686 -0 0 -0 -1.07264e-17
      6.3178 -0.430704 -0 0 -0 -1.91593e-17 7.38567 -1.4624 -0 0 -0
       -2.45986e-17 9.22288 -0.888062 -0 0 -0 -2.82008e-17 10.2464
      1.53133 -0 0 -0 -1.25459e-17 4.64811 0.694664 -0 0 -0
      -1.75924e-17 4.68175 0.4096 -0 0 -0 -1.9515e-17 4.6908
      0.74295 -0 0 -0 1.19736 -7.38744 7.31601 -0 0 -0 12.5803
      3.15559 -3.17425 -0 0 -0 28.3639 0.0580795 0.133078 -0 0 -0
      23.5335 0.0481878 0.110415 -0 0 -0 -1.19736 -7.38743 7.31591
      -0 0 -0 -12.5803 3.15559 -3.17425 -0 0 -0 -28.3639 0.0580795
      0.133078 -0 0 -0 -23.5335 0.0481876 0.110415 -0 0 -0 8.96978
      -4.08401 1.97395 -0 0 -0 -0.827718 -37.6515 -0.63625 -0 0 -0
      -2.25899 -39.2341 -5.87393 -0 0 -0 0.756579 -9.59872 12.1348
      -0 0 -0 -8.96978 -4.08401 1.97395 -0 0 -0 0.827718 -37.6515
      -0.63625 -0 0 -0 2.25899 -39.2341 -5.87393 -0 0 -0 -0.756579
      -9.59872 12.1348 -0 0 -0
```

付録 B

Intel RealSense D415 のスペック

表 B.1: Intel RealSense D415 のスペック

使用環境	屋内/屋外
深度技術	Active IR stereo(ローリングシャッター)
主要 Intel RealSense コンポーネント	Intel RealSense Vision Processor D4
	Intel RealSense module D410
深度センサ視野角 (水平 × 垂直 × 斜め)	69.4° × 42.5° × 77° (± 3°)
出力解像度 (DepthStream)	最大 1280 × 720
出力フレームレート (DepthStream)	最大 90 fps
最小深度距離 (Min-Z)	0.3 m
シャッタータイプ	ローリングシャッター
最大レンジ	約 10 m(較正，背景，照度状況による)
解像度およびフレームレート (RGB センサ)	1920 × 1080@30 fps
RGB センサ視野角 (水平 × 垂直 × 斜め)	69.4° × 42.5° × 77° (± 3°)
本体寸法 (長さ × 奥行き × 高さ)	99 mm × 20 mm × 23 mm
コネクタ	USB 3.0 Type-C
取付機構	1 × 1/4-20 UNC ネジ穴、2 × M3 ネジ穴
測定原理	アクティブステレオ法

付録 C

オイラー角形式

オイラー角 [8] とは，三次元ユークリッド空間中の 2 つの直交座標系の関係を表現する方法の一つである．レインハルト・オイラーによって考案されたもので，剛体に固定された座標系を考えることで剛体の姿勢を表すことができる．

オイラー角は 3 つの角度の組で表され，オイラー角で繋がれる 2 つの座標系のうち地上に固定された座標系を (x,y,z)，剛体に固定された座標系を (X,Y,Z) で表すとした場合，以下のような手順でオイラー角を求めることができる．

1. z 軸と Z 軸のなす角度を β とする
2. β が 0° または 180° ではない場合には、x-y 平面と X-Y 平面は一つの直線で交わる。この交線を N とする
3. x 軸と交線 N のなす角度を α とし、X 軸と交線 N のなす角度を γ とする

この時 (α, β, γ) がオイラー角である．オイラー角は座標軸まわりの回転を繰り返すことで表すこともできる．