

NAIST-IS-MT1651002

修士論文

RGB カメラのステレオ視による 3 次元手形状の復元

尼子 琢朗

2018 年 3 月 10 日

奈良先端科学技術大学院大学
情報科学研究科 情報科学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に
修士（工学）授与の要件として提出した修士論文である。

尼子 琢朗

審査委員：

小笠原 司 教授	（主指導教員）
杉本 謙二 教授	（副指導教員）
高松 淳 准教授	（副指導教員）
丁 明 助教	（副指導教員）

RGB カメラのステレオ視による 3 次元手形状の復元*

尼子 琢朗

内容梗概

我々人間の手は日常生活を送る上で重要な運動器官である。その機能のひとつに、モノを操作する把持動作が挙げられる。これは、ジェスチャーのような動作とは異なり、手と物体のインタラクションが求められる。把持動作に関しては様々な研究が行われてきたが、センサグローブのような手の動きを妨げる計測方法では自然な動作とは言えない。したがって、新たに手を拘束しない計測手法を考える。

本研究では、より自然な形で人の手形状を推定するため、RGB カメラを用いたステレオ視の計測システムを提案し、カメラ画像から 3 次元空間中の手形状を復元することを目的とする。手形状の復元を、手のリンク構造の取得と三角測量法を用いた深さ方向の推定に分けて考える。単眼の RGB カメラが記録したデータから 2 次元の手のリンク構造を推定し、左右のカメラから見た時の対応点を基に三角測量法を用いて 3 次元形状に復元する。また、本研究では計測誤差の対策として、2 次元のリンク構造の誤差を低減するために非線形最適化法のひとつである Nelder-Mead 法を実装する。

実験では、構築した計測システムによって推定される、3 次元空間中に復元された手の関節位置について評価する。手を静止させている状態で計測し、真値との誤差を比較することで誤差の要因について検証する。真値はモーションキャプチャによって計測する。

キーワード

手形状, 3 次元復元, RGB カメラ, ステレオビジョン, Nelder-Mead 法

*奈良先端科学技術大学院大学 情報科学研究科 情報科学専攻 修士論文, NAIST-IS-MT1651002, 2018 年 3 月 10 日.

Reconstruction of 3D Hand Shape by Stereo Vision of RGB Camera*

Takuro Amako

Abstract

Human hands are important exercise organs to live everyday life. One of the functions is a grasping motion for manipulating objects. Unlike gesture-like motion, grasping motion requires hand and object interaction. Various researches have been conducted on the grasping motion, but since it can not be said that it is a natural motion in the measurement method that hinders the movement of the hand like a sensor glove, consider a measurement method that does not newly hold the hand.

In this research, we propose a stereo vision measurement system using RGB camera to estimate human hand shape in a more natural form and aim to restore the hand shape in three dimensional space from the camera image. Restoration of hand shape is divided into acquisition of link structure of hand and estimation in depth direction using triangulation method. Estimate the link structure of the two-dimensional hand from the data recorded by the monocular RGB camera and restore it to the 3-dimensional shape by triangulation method based on corresponding points when viewed from the left and right cameras. In addition, in this research, as a countermeasure against measurement error, Nelder-Mead method which is one of nonlinear optimization method is implemented to reduce error of 2-dimensional link structure.

In the experiment, we evaluate the joint position of the hand restored in the 3-dimensional space, estimated by the constructed measurement system. Mea-

*Master's Thesis, Department of Information Science, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT1651002, March 10, 2018.

sure with the hand in a stationary state, and verify the cause of the error by comparing the error with the true value. We measure the true value by motion capture.

Keywords:

Hand Shape, 3D Reconstruction, RGB Camera, Stereo Vision, Nelder-Mead Simplex Method

目次

図目次	vi
-----	----

表目次	viii
-----	------

第 1 章	はじめに	1
1.1	研究背景	1
1.2	関連研究	2
1.3	研究目的	2
1.4	本論文の構成	3
第 2 章	手形状の推定	4
2.1	RGB 画像による 2 次元リンク構造の取得	4
2.2	手形状の 3 次元復元	7
2.3	Nelder-Mead 法	8
第 3 章	手形状計測システム	10
3.1	RGB カメラ	10
3.1.1	ステレオキャリブレーション	11
3.1.2	三角測量の確認	13
3.2	システム構成	15
第 4 章	手形状復元実験	18
4.1	実験環境	18
4.2	実験条件	22
4.3	実験結果	24
4.4	考察	32
4.4.1	推定誤差に対する考察	32
4.4.2	定常誤差に対する考察	32
4.4.3	手の位置を変更した場合の手形状推定	34

第 5 章 おわりに	40
謝辞	42
参考文献	43

図目次

2.1	Estimated hand feature points	5
2.2	Estimated pose feature points	6
2.3	Triangulation by two viewpoints	7
3.1	Appearance of C270	10
3.2	Pedestal for fixing RGB cameras	11
3.3	Stereo vision setup	12
3.4	Calibrated stereo view	12
3.5	Confirmation of triangulation	14
3.6	Detection of the circle from stereo vision	15
3.7	Proposed measurement system	15
3.8	System configuration for measurement system	17
4.1	Measurement setup	19
4.2	Configuration of the experimental environment	20
4.3	Appearance of the experimental environment	21
4.4	Motion capture marker on the hand	22
4.5	Arrangement each link	23
4.6	OpenPose estimation from each view and reconstructed hand shape	25
4.7	Mean value of error in each joint	26
4.8	Joints annotation of human hand	26
4.9	Error in thumb	27
4.10	Error in index finger	28
4.11	Error in middle finger	29
4.12	Error in ring finger	30
4.13	Error in little finger	31
4.14	Data in the frame with the largest estimation error	33

4.15	Difference in the x axis	35
4.16	Difference in the y axis	36
4.17	Difference in the z axis	37
4.18	OpenPose estimation from each view and reconstructed hand shape in the case where the hand position is changed	38
4.19	Mean value of error in each joint in the case where the hand position is changed	39

表目次

3.1	Specification of RGB camera	11
4.1	Specification of V120:Duo	21
4.2	Each link length	23
4.3	Estimation error in the experiment	24
4.4	Estimation error in thumb	27
4.5	Estimation error in index finger	28
4.6	Estimation error in middle finger	29
4.7	Estimation error in ring finger	30
4.8	Estimation error in little finger	31
4.9	Difference MCP joint position in the x axis	35
4.10	Difference MCP joint position in the y axis	36
4.11	Difference MCP joint position in the z axis	37
4.12	Estimation error in the case where the hand position is changed	39

第 1 章 はじめに

1.1 研究背景

我々人間の手は日常生活を送る上で重要な運動器官である。また、手の機能は手根骨や中手骨、指骨と呼ばれる 27 の骨格と、関節や腱、筋肉、皮膚などの複雑な構造によって実現する事が可能な、多様な動きによって支えられている [1] [2]。手がおもつ重要な機能のひとつとして、モノを掴み操作する把持動作がある。この動作に関して、人間が行う把持動作がどのように変化するか、また把持時の手形状についてどのような分類ができるのかといった研究は古くから行われている [3] [4]。

Napier は人間が把持の型を決定する要素は把持対象の形状や大きさによってではなく、把持を行う目的により決まるとし、把持を行う際には「力」と「精緻性」のどちらがより必要とされるかによって握力把持、又は精密把持のどちらかが選ばれているとした [3]。また、Cutkosky は Napier の理論を基に、工場作業員の手の動かし方を観察し、工場で利用される把持形態を 16 種類に細分化した [4]。

近年では、把持動作について人の手形状や把持物体に加える力を計測するといった研究も行われている。幸坂ら [5] は利き手と非利き手の役割の違いを調査するためにウェアラブルカメラと筋電位、加速度センサを複合したセンサを利用し、日常生活の場面の記録と分析を試みた。Garcia-Hernand ら [6] は頭部に RGB-D センサ、手にセンサグローブを取り付け、日常生活における手形状を分類し、データセットを作成した。また、Ju ら [7] は手にかかる圧力と関節角度、筋電位を計測できるセンサグローブによって 10 種類の把持動作に関して分類した。このように、人の把持動作については多方面で研究が盛んに行われている。

ここで、把持以外の手の機能であるジェスチャーと把持動作について考えると、ジェスチャーは手形状と腕の動きの変化に着目するのに対し、把持動作では物体の形状や手の把持形状、把持位置のように手と物体の動作に着目する。したがって、ジェスチャーの計測では手形状の違いを計測するため、センサグローブなどが有効であるが、把持動作では手と物体の動きを阻害しない計測が有効であると考えられる。

1.2 関連研究

手と物体の動きを阻害しない計測について，ウェアラブルデバイスを用いた研究が行われている．山崎らは RGB-D カメラを頭部に取り付け，1 人称視点の動画からポイントクラウドを抽出し，手形状の推定と手動作を認識している [8]．また，伊藤らは筋電位計測装置の Myo とハンドジェスチャインタフェースの Leap Motion を統合したセンサを上腕に取り付け，日常生活における手の動作について上肢機能の評価方法である SHAP を対象として手動作の認識を行なっている．また，非接触による計測としては Joo ら [10] は 480 台の RGB センサをドーム状に配置し，人のボーンを推定するソーシャルモーションキャプチャを提案している．Simon ら [11] は，Joo らの研究の延長として 1 台の RGB センサのみで 2 次元の手形状を推定している．

先行研究では，把持動作における手動作について，手と物体を阻害しない計測が提案されている．本研究では，完全に自由な把持動作計測システムを最終的な目標として，手形状に限った計測システムを構築する．

1.3 研究目的

本研究では，RGB カメラを用いて手に対して非接触な計測システムを構築し，3 次元空間中の手形状の復元を目的とする．本システムでは RGB カメラによるステレオ視をおこなうため，専用の台座に取り付けられた RGB カメラをステレオカメラとして扱い，計測をおこなう．

手形状の復元は，手のリンク構造の取得と三角測量法を用いた深さ方向の推定に分けて考える．単眼の RGB カメラが記録したデータから 2 次元の手のリンク構造を推定し，左右のカメラから見た時の対応点を基に三角測量法を実施する．また，本研究では計測誤差の対策として，3 次元復元の誤差を低減するために非線形最適化法のひとつである Nelder-Mead 法を実装する．

実験では，構築した計測システムを用いて，3 次元空間中に復元された手の関節位置を推定する．具体的には，手を静止させている場合で真値との誤差を比較する．加えて，誤差の要因について検証する．真値はモーションキャプチャを用いて計測される 3 次元座標である．

1.4 本論文の構成

本論文の構成は以下の通りである. 第 2 章では, RGB 画像から手形状を推定するまでの手法と推定誤差の低減について提案し, 第 3 章にて本研究で使用する計測システムを構築する. 第 4 章では計測システムを用いて手形状復元の実験と考察を行う. 最後に, 第 5 章で本研究のまとめと今後の課題について述べる.

第 2 章 手形状の推定

本章では，手形状の推定について，3 次元空間中の手の位置と各関節位置を推定する手法について述べる．具体的には，以下の 3 点により実現を図る．

- 単眼の RGB 画像からカメラ座標系における 2 次元の手のリンク構造を推定
- 水平ステレオ構築し，左右の視点からリンク構造を取得
- 三角測量法に基づいて 3 次元リンク構造に復元

2.1 RGB 画像による 2 次元リンク構造の取得

人の手の動作を RGB カメラなどを用いて非接触で計測する場合，画像から人の手の部分を抽出し，手の関節位置や指先位置を推定する必要がある．本研究では，Simon ら [11] によって提案され，GitHub 上に公開されている OpenPose [12] を用いる．OpenPose とは，単眼の RGB 画像から深層学習を用いて人を認識し，その関節などの特徴点を検出するアルゴリズムである．推定対象は人の姿勢，顔，手であり，各特徴点を [" x 軸方向のピクセル値", " y 軸方向のピクセル値", "その点における尤度"] の順に出力する．それぞれの値の範囲は [0-画像の幅]，[0-画像の高さ]，[0-1] である．OpenPose では，オプションとして推定する項目を選択出来るが，顔と手については姿勢の推定と併せて実行する必要があったため，本研究では姿勢と手の 2 つの推定を行う．これらの特徴点の出力イメージを Fig. 2.1 - 2.2 に示す．OpenPose による手の推定では，手首を起点として各指の関節と指先を含む合計 21 点の 2 次元リンク構造が出力される．

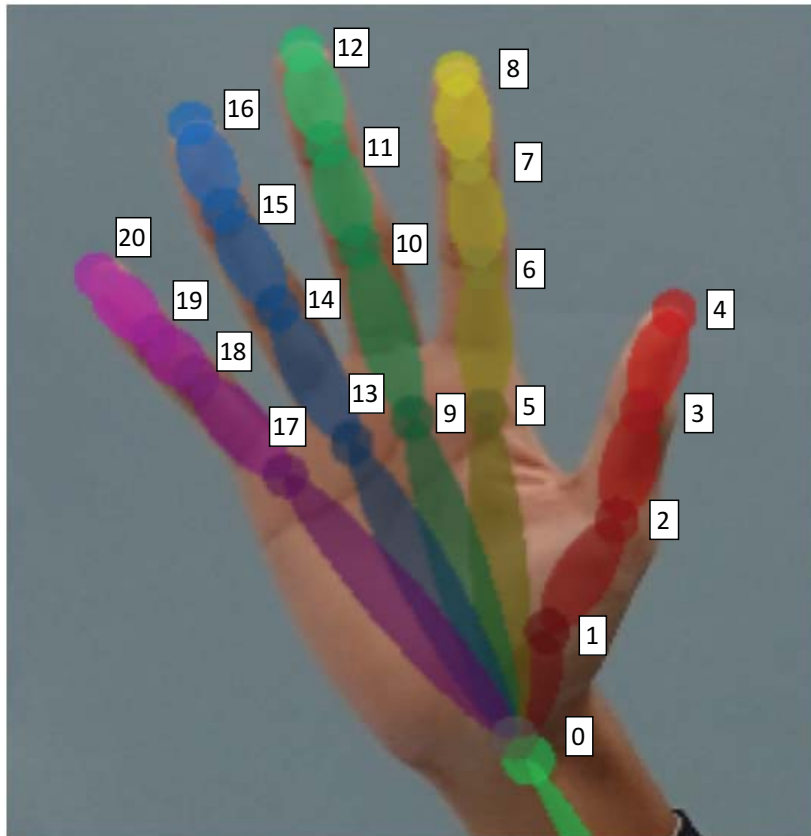


Fig.2.1 Estimated hand feature points

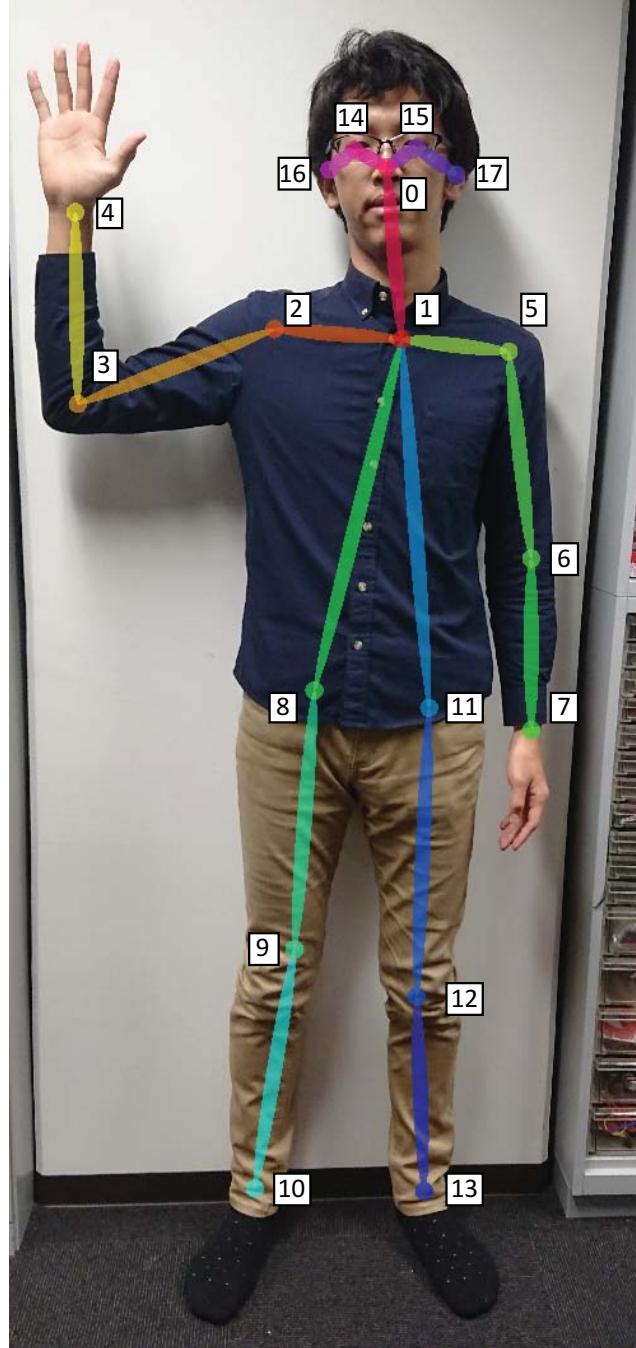


Fig.2.2 Estimated pose feature points

2.2 手形状の 3 次元復元

2 次元リンク構造を 3 次元に復元するため, RGB カメラ 2 台を用いてステレオキャリブレーションを行い, ステレオカメラを構築する. それぞれのカメラから取得する画像を OpenPose にかけることで, 左右の視点から見た手の 2 次元リンク構造を得る. そして, 対応する特徴点ごとに三角測量法を適用することで, 空間中における特徴点の位置を推定し, もとのリンク構造にしたがって手形状を 3 次元に復元する.

2 視点による三角測量を Fig. 2.3 に示す. それぞれ O_l, O_r を基準とする焦点距離 f の 2 台のカメラについて, 中心間距離 (基線長) が b で X, x_l, x_r の軸方向が揃うような水平ステレオの場合, 次のように 3 次元点 $P(X, Y, Z)$ を求めることができる.

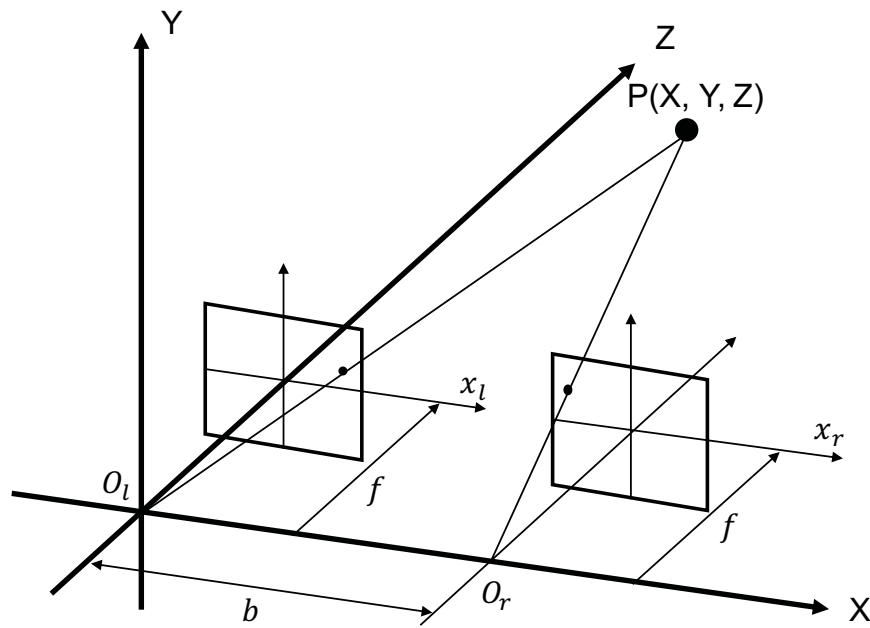


Fig.2.3 Triangulation by two viewpoints

$$X = \frac{x_l}{x_l - x_r} b \quad (2.2.1)$$

$$Y = \frac{y_l}{x_l - x_r} b = \left(\frac{y_r}{x_l - x_r} b \right) \quad (2.2.2)$$

$$Z = \frac{f}{x_l - x_r} b \quad (2.2.3)$$

2.3 Nelder-Mead 法

本研究の手法では、3次元復元された手形状に関して、推定誤差の要因となり得る項目が2つ考えられる。1つはステレオ視の精度である。これは2.2節で述べたキャリブレーションの精度に直結する。カメラの位置関係が変わる度にキャリブレーションを実施する必要があることに注意しなければならない。もう1つはOpenPoseの推定値である。手形状の3次元復元はこの推定値を基に実施するため、推定誤差への対策が要求される。

手形状の誤差を低減するために、2次元リンク構造に対してNelder-Mead法[14]を実装する。この手法は滑降シンプレックス法やアメーバ法とも呼ばれており、非線形最適化法的一种として多次元非線形関数の最小値を探索する。一般的な非線形最適化手法では、最小化する目的関数の微分情報を使用するため、微分不可能な場合や目的関数自体が不明の場合では最適化を実行できない。しかし、Nelder-Mead法はそのような情報がなくても非線形最適化を実行できるという特徴を持つ。一方で、収束に時間がかかってしまうという問題がある。

本研究では、各フレームでの21点のそれぞれの特徴点に対して以下の関数 $f(i, j)$ (2.3.1)を目的関数とする。ここで、 w_1, w_2, w_3 はそれぞれの項の重み、 i は記録したデータのフレーム数、 j はOpenPoseの特徴点($0 \leq j \leq 20$)、 P_* ($*$: *Left, Right*)は i 番目のフレームにおける j 番目の特徴点のカメラ座標とする。 $Proj_*$ ($*$: *Left, Right*)は三角測量によって算出された3次元座標をOpevCVの`cv2.projectPoints`関数を用いて左右のカメラ座標系にそれぞれ投影したときの座標とする。

また，3次元復元で推定された21の特徴点から，Fig. 2.1に示す各リンクに対応する2つの特徴点に関してユークリッド距離求め，その値をリンク長 $length$ とする． l を事前に計測したリンク長とする．

$$\begin{aligned}
 f(i, j) = & w_1 (Proj_L(i, j) - P_L(i, j))^2 \\
 & + w_2 (Proj_R(i, j) - P_R(i, j))^2 \\
 & + w_3 (length_{ij} - l_j)^2
 \end{aligned} \tag{2.3.1}$$

第 3 章 手形状計測システム

本研究では，人の手形状を計測するために，RGB カメラである logicool 社の HD Webcam C270 を使用する．本章では，ステレオカメラを構成するための専用の台座を 3D プリンタを用いて作成した．その後，台座に取り付けた RGB カメラについてキャリブレーションを行い，ステレオ視の確認と三角測量について検証した．

3.1 RGB カメラ

手形状を計測するセンサとして RGB カメラである logicool 社製の HD Webcam C270 [15] を使用する．C270 の外観を Fig. 3.1 に，仕様を Table 3.1 に示す．本研究では，C270 を 2 台使用して手形状推定を行う．



Fig.3.1 Appearance of C270

Table.3.1 Specification of RGB camera

Size	Width	70 mm
	Height	32 mm
	Depth	69 mm
Weight		71 g
FOV		60 deg
Resolution		640 × 480
Frame rate		30 fps
Connection		USB 2.0

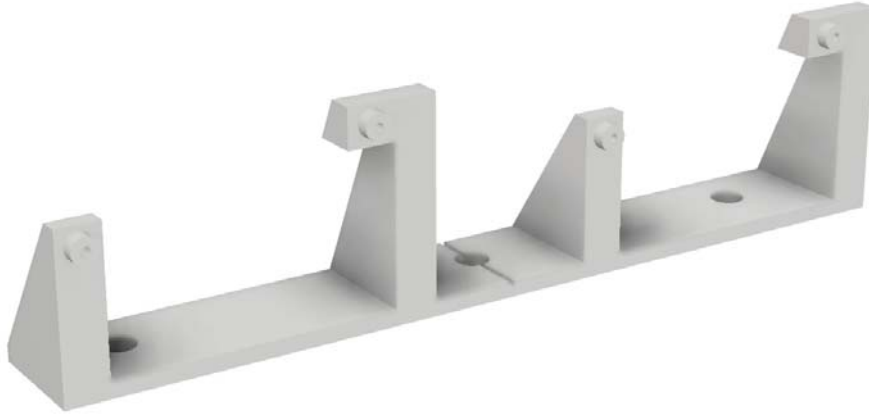


Fig.3.2 Pedestal for fixing RGB cameras

3.1.1 ステレオキャリブレーション

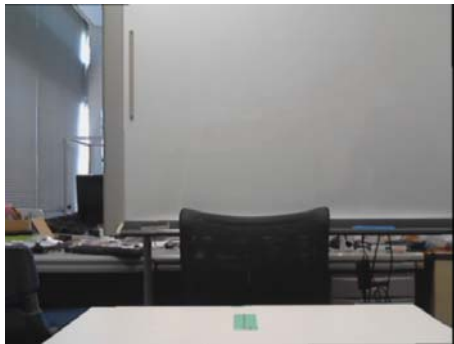
ステレオ視のためには RGB カメラを固定する必要がある. しかし, C270 は Fig. 3.1 の状態では固定させるための機構がないため, カメラ全体ではなく, 外側のフレームを分解して内部の基板を固定することにした. そこで, 3D プリンタを用いて RGB カメラを固定する専用の台座を ABS 樹脂で作成した. その外観を Fig. 3.2 に示す. この台座は, 設置するカメラ同士の基線長が水平方向に 100 [mm] だけ離れるように設計されている.

台座に 2 台の RGB カメラを固定した状態を Fig. 3.3 に示す. この状態で Zhang の方法 [13] によるステレオキャリブレーションを実施した. 各パラメータの算出には OpenCV の `cv2.stereoCalibrate` 関数と `cv2.stereoRectify` 関数を使用した. 得られた左右の内部パラメータ行列 C , 歪み係数行列 d , 外部パラメータ行列 $[R|t]$, および投影行列 P を以下に示す. なお, 本章では左カメラの光学中心をワールド座標系の原点としている.

最後に, キャリブレーションにより得られたパラメータをそれぞれのカメラに適用した. そのときの左右の視点を Fig. 3.4 に示す.



Fig.3.3 Stereo vision setup



(a) Left view



(b) Right view

Fig.3.4 Calibrated stereo view

$$C_{left} = \begin{bmatrix} 804.4 & 0.000 & 340.4 \\ 0.000 & 804.0 & 241.5 \\ 0.000 & 0.000 & 1.000 \end{bmatrix}$$

$$C_{right} = \begin{bmatrix} 814.3 & 0.000 & 316.9 \\ 0.000 & 813.5 & 258.5 \\ 0.000 & 0.000 & 1.000 \end{bmatrix}$$

$$d_{left} = \begin{bmatrix} 0.04146 & -0.2344 & -0.001637 & 0.001984 & 0.5015 \end{bmatrix}$$

$$d_{right} = \begin{bmatrix} -0.001311 & 1.029 & 0.0007673 & 0.0005446 & -3.298 \end{bmatrix}$$

$$[R|t] = \begin{bmatrix} 0.9999 & -0.005324 & 0.001388 & -98.90 \\ 0.005351 & 0.9998 & -0.02033 & -0.9487 \\ -0.001280 & 0.02034 & 0.9998 & 0.5498 \end{bmatrix}$$

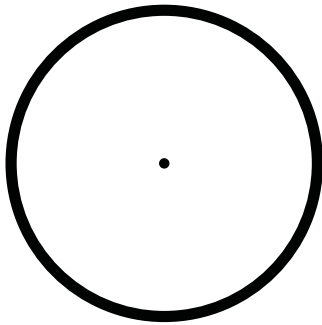
$$P_{left} = \begin{bmatrix} 804.0 & 0.000 & 333.8 & 0.000 \\ 0.000 & 804.0 & 249.7 & 0.000 \\ 0.000 & 0.000 & 1.000 & 0.000 \end{bmatrix}$$

$$P_{right} = \begin{bmatrix} 804.0 & 0.000 & 333.8 & -79510 \\ 0.000 & 804.0 & 249.7 & 0.000 \\ 0.000 & 0.000 & 1.000 & 0.000 \end{bmatrix}$$

3.1.2 三角測量の確認

次に、このステレオカメラによって3次元位置の推定が可能であることを検証する。ここでは、Fig. 3.5 の環境で奥行き方向に対しての距離を推定する。具体的には、RGB カメラから 1200 [mm] の距離にカメラと平行に配置されている Fig. 3.5(a) の円を認識し、左右のカメラから得られた円の中心座標 (単位: pixel) を OpenCV の `cv2.triangulatePoints` 関数に代入する。

計測時の様子を Fig. 3.6 に示す。検出された円の中心座標は左が [272, 350]、右が [206, 350] であった。この値をもとに三角測量を行うと、1204 [mm] という結果が得られ、真値に対して +0.33 % 誤差で奥行きを推定できることが確認できた。



(a) Circle

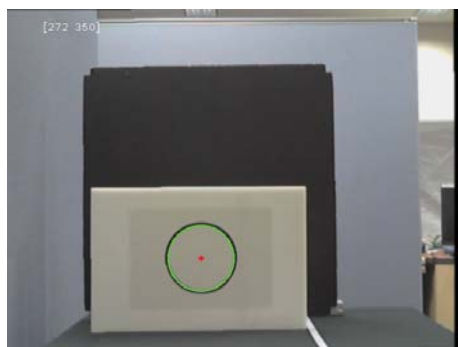


(b) Distance to measure

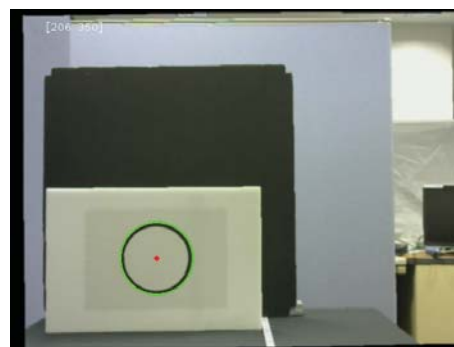


(c) Distance from RGB cameras

Fig.3.5 Confirmation of triangulation



(a) Left view



(b) Right view

Fig.3.6 Detection of the circle from stereo vision

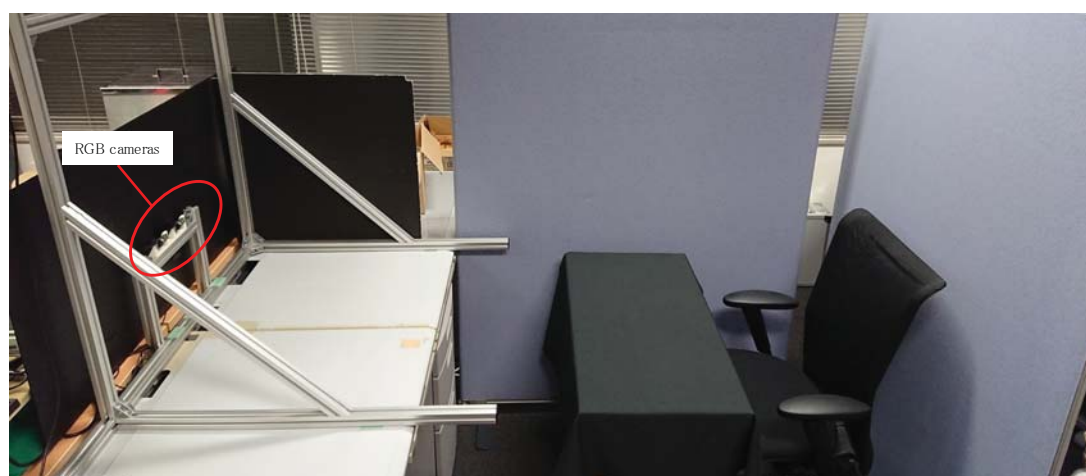


Fig.3.7 Proposed measurement system

3.2 システム構成

本研究で構築した計測システムについて、計測環境全体を Fig. 3.7 に示す. なお, RGB カメラの配置については Fig. 3.3 と同様である. このシステムでは, 固定されたフレームからカメラ前方にある人の手形状を推定する.

前節で述べた手形状計測システムのソフトウェア構成について説明する. 本シス

テムのソフトウェア構成は Fig. 3.8 に示す通りである. 本研究では, 配置の異なる 2 台のカメラから同時にデータを記録する必要がある. しかし, 記録を開始するタイミングは各デバイスにアクセスする時間に左右されてしまうため, わずかに誤差が生じてしまう. これを回避するため, UNIX 時間を参照することでデータの同期を実現する.

システム時間の取得とデータの記録は ROS (Robot Operating System) [16] を使用する. ROS はロボットソフトウェアプラットフォームの一種であり, 一般的には多関節システムや自律制御システムなどで, 多数のパラメータを同時に処理するシステムに用いられる. ROS ではノードと呼ばれるプロセスの間で通信が行われているのだが, ハードウェア上でもプロセス間通信を自由に行うことができる. 本研究ではロボットを使用しないが, ROS の特徴である通信の容易さに着目し, これをデータの同期に活用する.

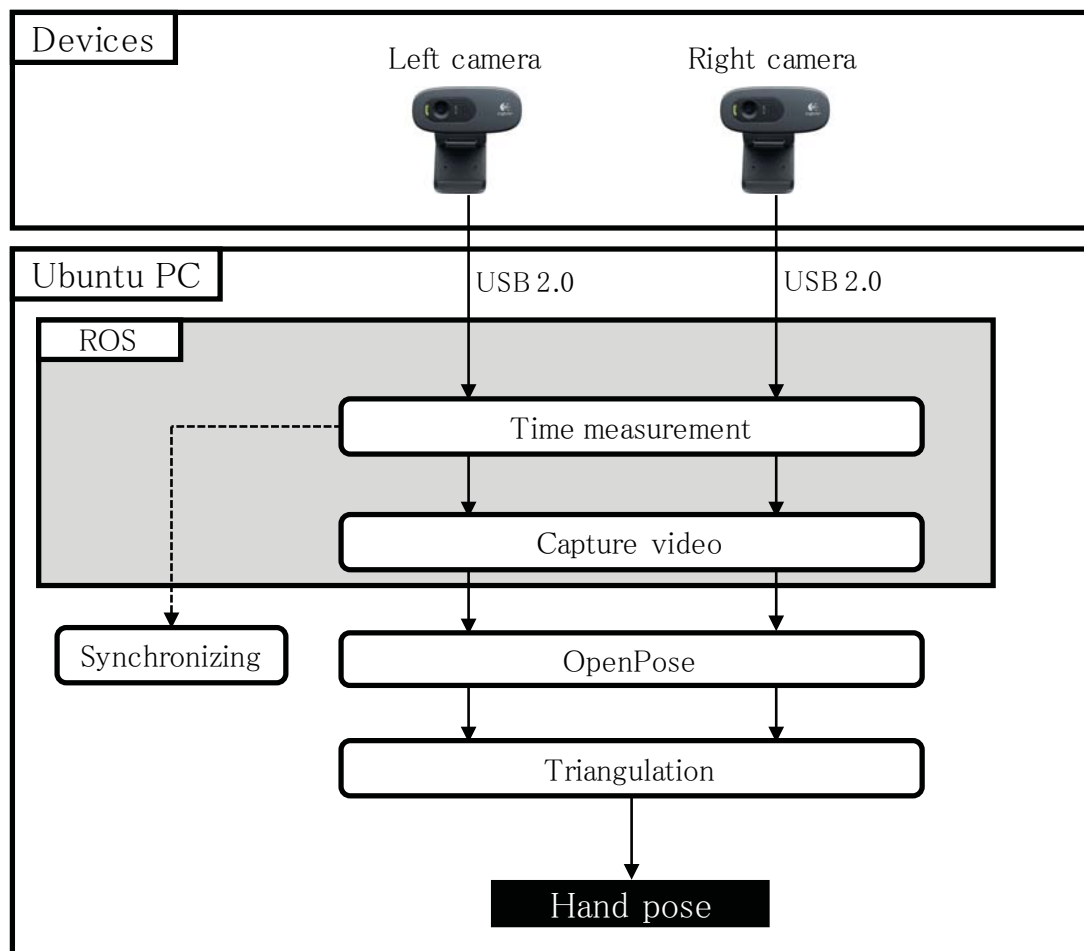


Fig.3.8 System configuration for measurement system

第 4 章 手形状復元実験

本章では、本研究で提案した計測システムによる手形状復元の実験をおこなう。実験は、モーションキャプチャシステムによる計測結果を真値として扱い、3次元復元された手形状の各特徴点を比較し、計測精度について評価することを目的とする。

4.1 実験環境

第 3 章で提案した計測システムに、真値を取得するため OptiTrack 社製のモーションキャプチャカメラ OptiTrack V120:Duo [17] を導入した。V120:Duo の仕様を Table. 4.1 に示す。また、計測装置の外観を Fig. 4.1 に示す。

実験では全ての計測を同時におこなうので、Fig. 3.8 の構成に V120:Duo を追加する必要がある。しかし、OptiTrack 社製のオブジェクトトラッキングソフトウェア Motive [18] が対応している OS は Windows 64-bit のみである。そこで、Fig. 4.2 のように Ubuntu PC と Windows PC を Ethernet で有線接続することによりデータ通信を実現した。他にも、それぞれのカメラから取得するデータは、カメラのフレームレートが異なるので、フレーム数を調整する必要がある。本実験ではレートの高い V120:Duo のデータを間引き、フレーム数を一致させる。このシステムではデータ同期のために UNIX 時間を参照していることから、ステレオカメラのデータを取得した時間に最も近い V120:Duo のデータを扱うこととする。

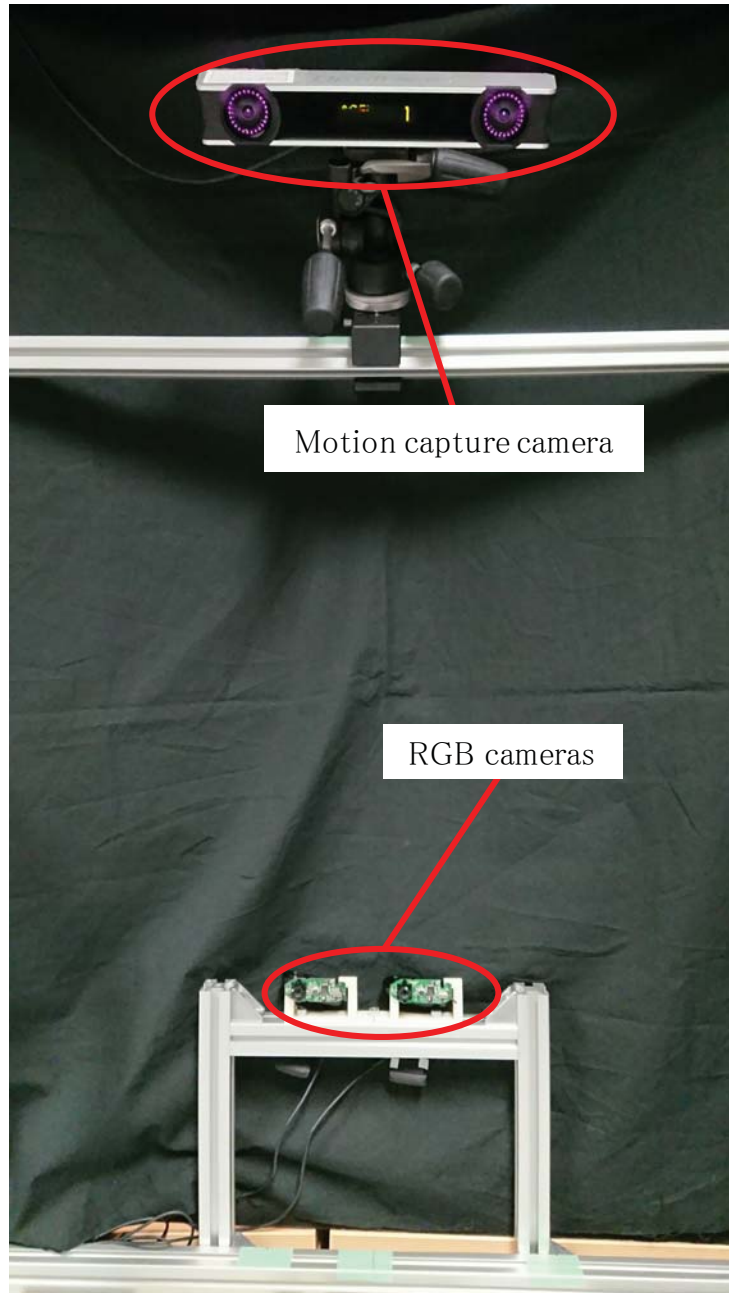


Fig.4.1 Measurement setup

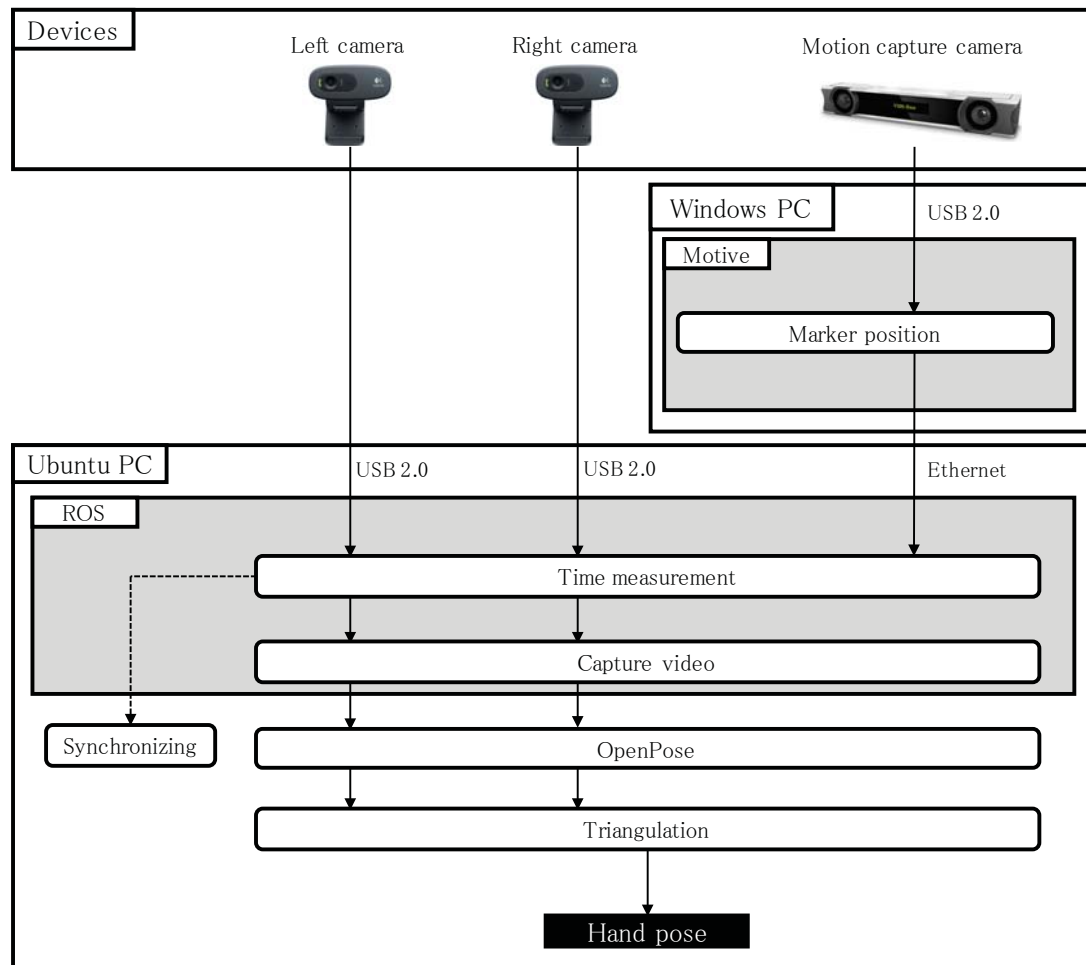


Fig.4.2 Configuration of the experimental environment

Table.4.1 Specification of V120:Duo

Size	Width	279.4 mm
	Height	40.6 mm
	Depth	50.8 mm
Weight		1,300 g
FOV	Horizontal	47 deg
	Vertical	43 deg
Connection		USB 2.0
Frame rate		120 fps



Fig.4.3 Appearance of the experimental environment

4.2 実験条件

4.1 節の装置を用いて, Fig. 4.3 に示す環境で実験をおこなった. 実験者は本論文の執筆者である. 計測装置に対して平行に手を開き, 静止させた状態を維持する動作をおこない, データを記録した.

モーションキャプチャで手形状を計測するため, Fig. 4.4 に示すように, OpenPose の特徴点に対応する位置に光学式反射マーカを貼り付けた. 加えて, Nelder-Mead 法で最小化する式 (2.3.1) に必要なリンク長 l_j をマーカ位置を参考にスケールで計測した. 各リンクの位置を Fig. 4.5 に, 計測したリンク長を Table. 4.2 に示す.



Fig.4.4 Motion capture marker on the hand

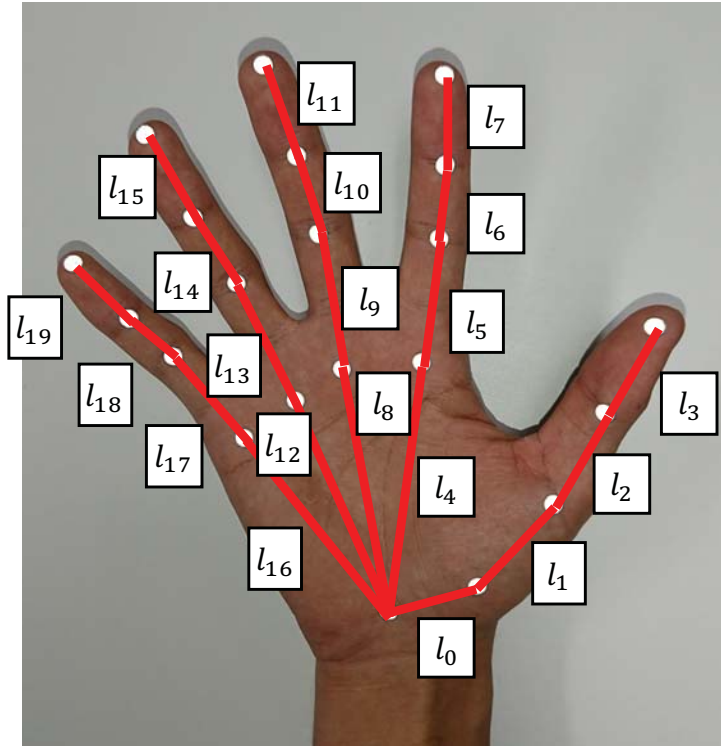


Fig.4.5 Arrangement each link

Table.4.2 Each link length

link	length [mm]	link	length [mm]	link	length [mm]
l_0	30.0	l_7	27.0	l_{14}	25.0
l_1	40.0	l_8	85.5	l_{15}	28.0
l_2	37.5	l_9	43.5	l_{16}	79.0
l_3	30.0	l_{10}	26.0	l_{17}	34.0
l_4	87.0	l_{11}	28.0	l_{18}	17.5
l_5	40.5	l_{12}	85.0	l_{19}	24.0
l_6	23.0	l_{13}	38.5		

4.3 実験結果

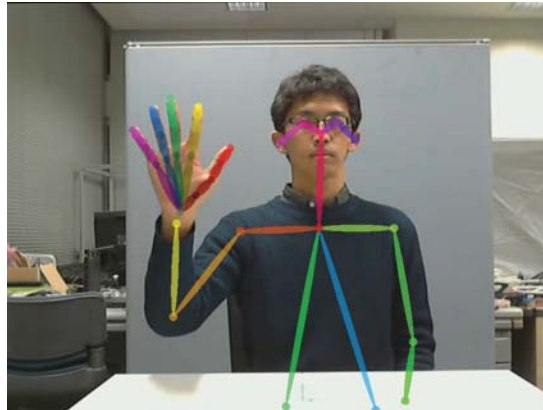
本実験で得られた推定結果を Fig. 4.6 に一部抜粋して示す. Fig. 4.6 (a), (b) はステレオカメラのそれぞれの視点から得られた OpenPose の推定値である. Fig. 4.6 (c) は 3 次元復元で得られた手形状を示しており, 手のひらと 5 指で色分けした. また, モーションキャプチャで計測されたマーカの位置を赤の円で描画した. この結果から, RGB 画像を用いて推定される手の 2 次元リンク構造が, ステレオ視によって 3 次元のリンク構造に復元されていることが分かる.

1 フレームごとに, モーションキャプチャで計測した各特徴点の 3 次元位置 p_i^M ($0 \leq i \leq 20$) を真値とし, ステレオカメラで推定した各特徴点の 3 次元位置 p_i^E とのユークリッド距離の平均 $\frac{1}{21} \sum_i \|p_i^M - p_i^E\|$ をそのフレームにおける推定誤差として計算した. 全フレームの時系列データを Fig. 4.7 に, 全フレームに対する推定誤差の平均と標準偏差を Table. 4.3 に示す. 計算の結果, データの前半部分では数値に大きな波があったが, 後半は一定の値に収束していることが分かる.

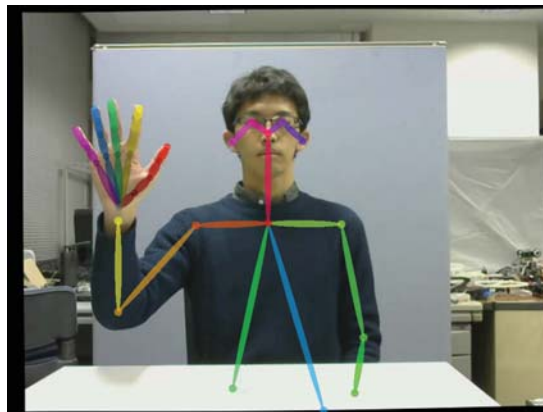
次に, 各指での推定誤差をユークリッド距離 $\|p_i^M - p_i^E\|$ で計算した. ここで, 各指間での比較を簡単にするため, 手の関節の名称を定める. 人の手の関節は Fig. 4.8 に示すように区別できることが知られており [19], 本研究もこれに則る. 加えて, 指先の特徴点を Fingertip とする. 各指における推定誤差について Fig. 4.9 - 4.13 に, 関節ごとの時系列平均と標準偏差を Table. 4.4 - 4.8 に示す. どの指においても, 指先の推定誤差が大きくなる傾向を確認できるため, 手形状の 3 次元復元において指先の特徴点の推定が困難であると分かる.

Table.4.3 Estimation error in the experiment

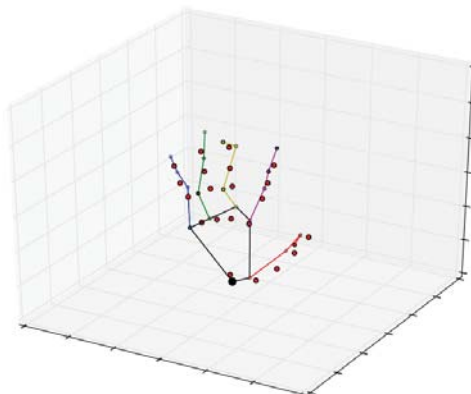
Mean error [mm]	Standard deviation
78.9	6.01



(a) OpenPose estimation (left)



(b) OpenPose estimation (right)



(c) Reconstructed hand and markers

Fig.4.6 OpenPose estimation from each view and reconstructed hand shape

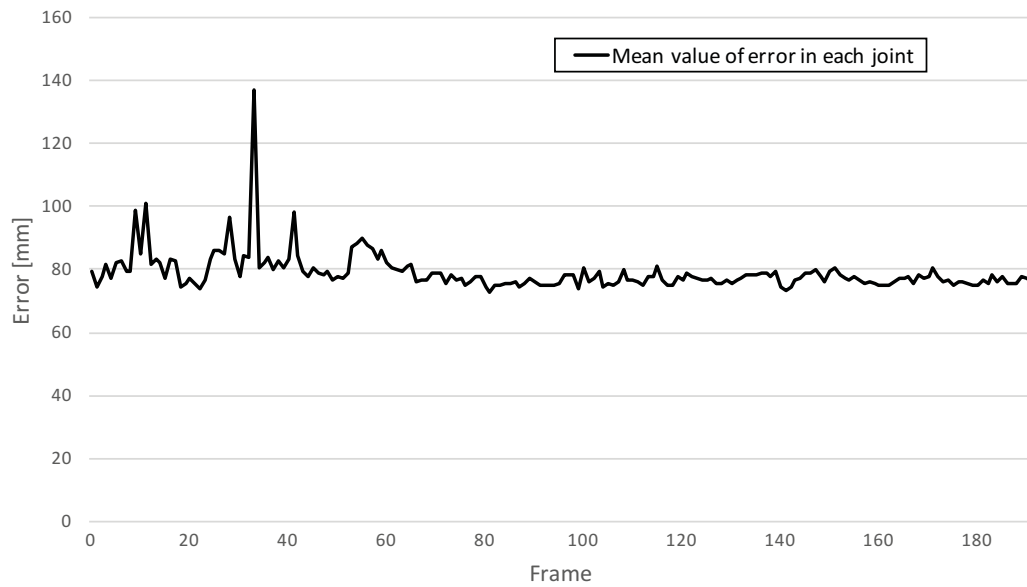


Fig.4.7 Mean value of error in each joint

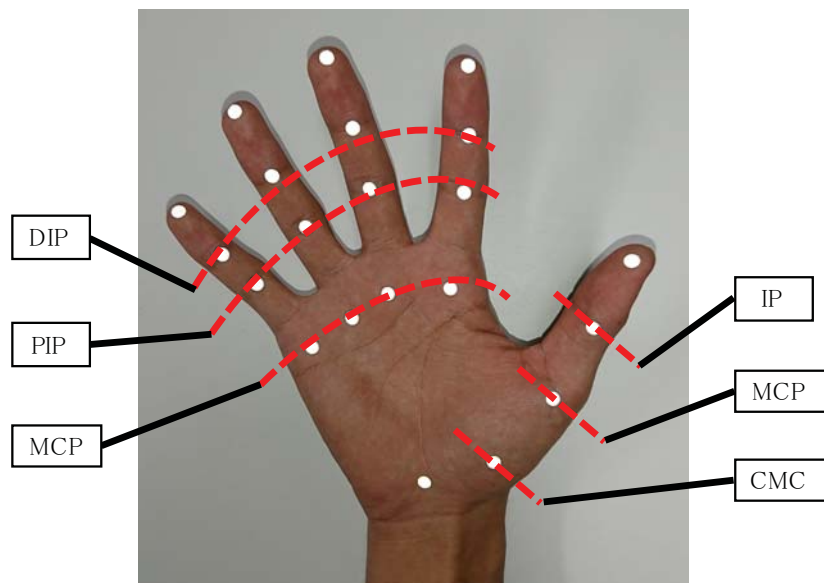


Fig.4.8 Joints annotation of human hand

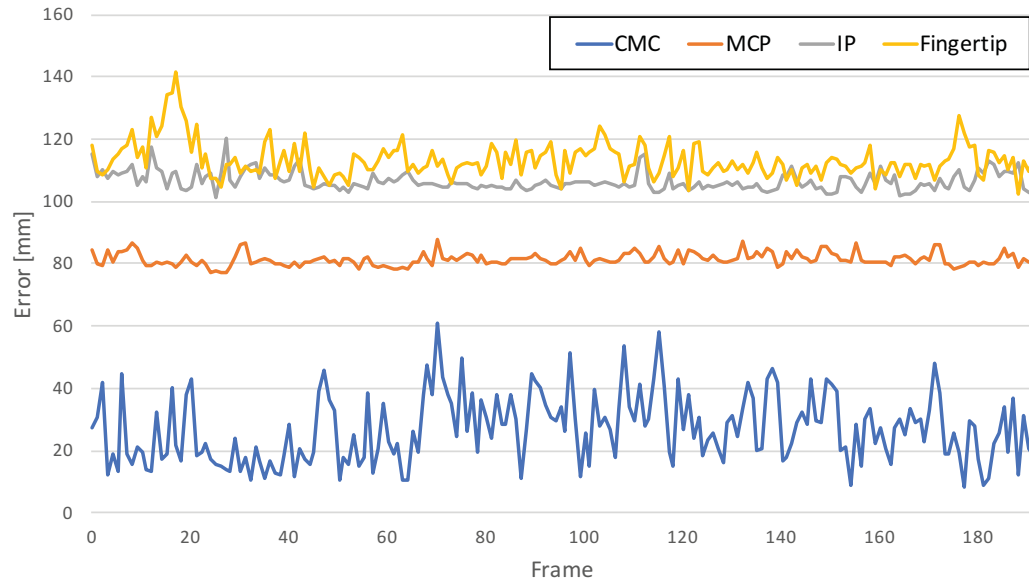


Fig.4.9 Error in thumb

Table.4.4 Estimation error in thumb

Joint	Mean error [mm]	Standard deviation
CMC	26.8	10.2
MCP	81.4	1.92
IP	106	2.70
Fingertip	113	4.48

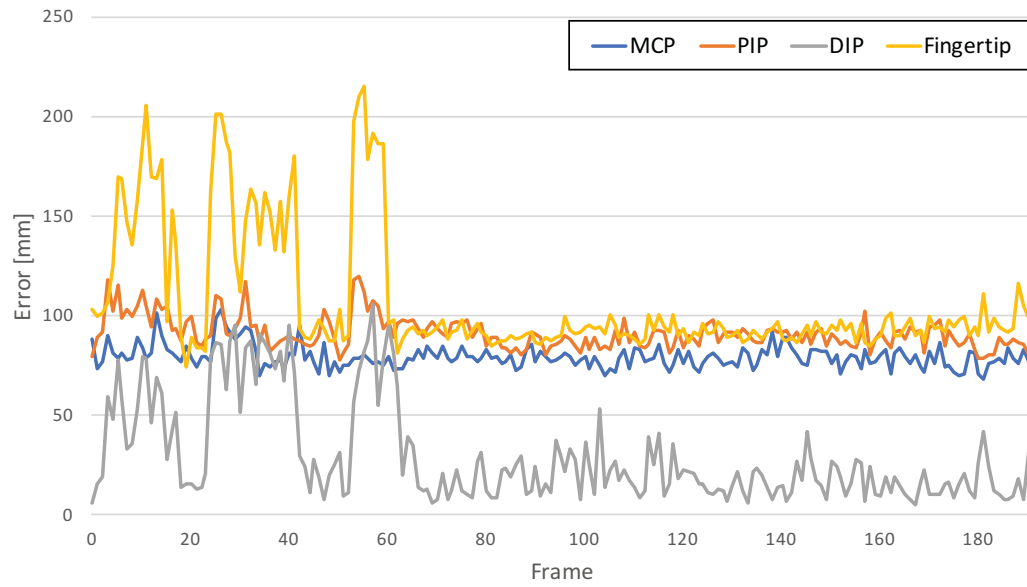


Fig.4.10 Error in index finger

Table.4.5 Estimation error in index finger

Joint	Mean error [mm]	Standard deviation
MCP	79.3	4.56
PIP	91.2	4.71
DIP	29.7	9.59
Fingertip	107	5.07

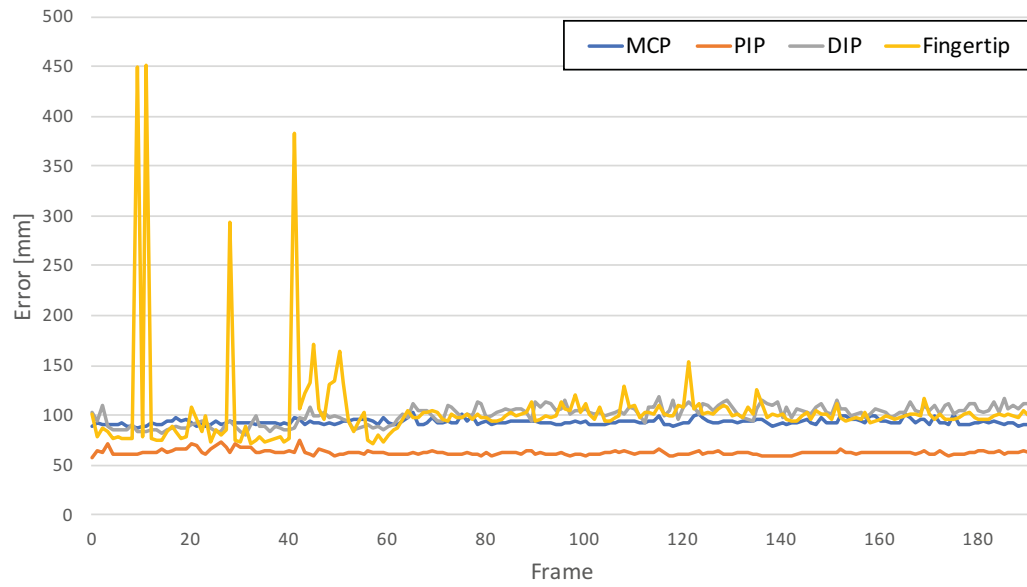


Fig.4.11 Error in middle finger

Table.4.6 Estimation error in middle finger

Joint	Mean error [mm]	Standard deviation
MCP	93.1	2.85
PIP	62.3	1.54
DIP	100	5.48
Fingertip	104	8.29

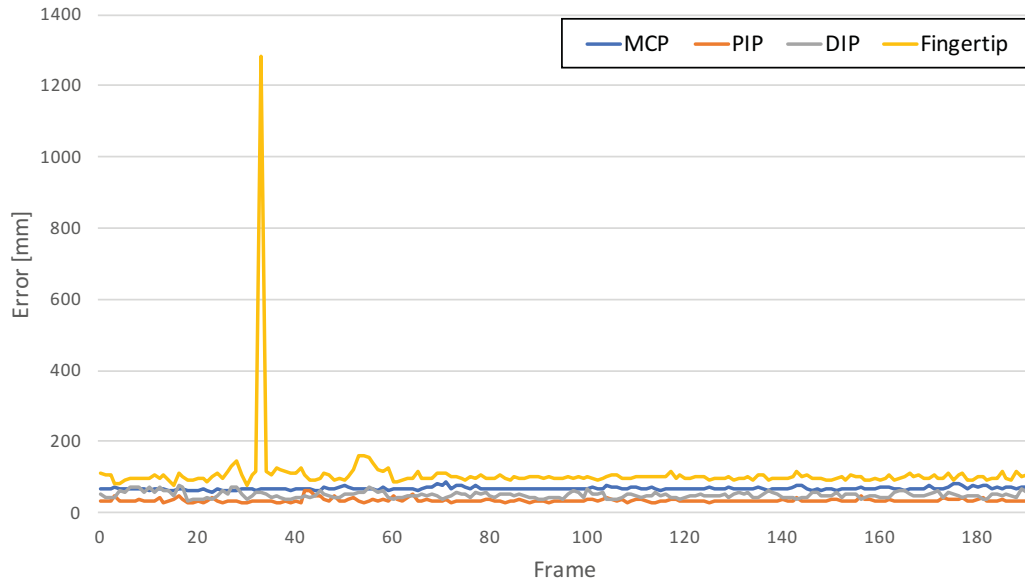


Fig.4.12 Error in ring finger

Table.4.7 Estimation error in ring finger

Joint	Mean error [mm]	Standard deviation
MCP	68.6	3.89
PIP	34.9	3.24
DIP	50.6	6.81
Fingertip	109	5.78

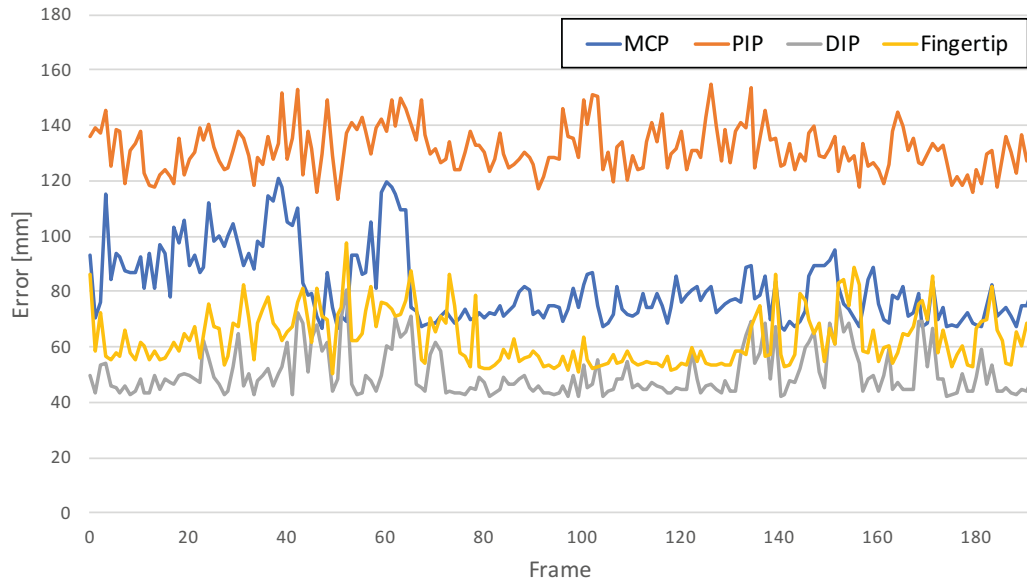


Fig.4.13 Error in little finger

Table.4.8 Estimation error in little finger

Joint	Mean error [mm]	Standard deviation
MCP	81.5	6.75
PIP	132	8.24
DIP	50.3	8.31
Fingertip	63.1	9.36

4.4 考察

実験の結果から、傾向としてデータの前半でいくつかの特徴点で大きな推定誤差が発生し、後のデータになるにつれて全体的に一定の値に収束していることが判明した。また、手を静止して計測するという実験条件であることから、収束する値は定常誤差であると捉えることができる。したがって、次の2点について考察する。加えて、手の位置によって推定結果にどのような影響が出るのかを実験し、検証する。

- データの前半で大きな推定誤差が発生した原因
- 3次元空間における定常誤差の大きさ

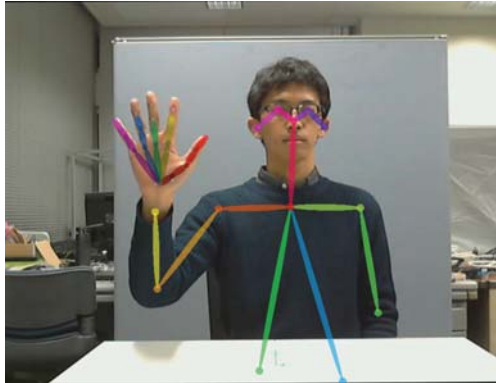
4.4.1 推定誤差に対する考察

推定誤差に関して、Fig. 4.7 から誤差が最大のときのフレームのデータを検証する。そのフレームでの OpenPose の推定結果と 3次元復元した手形状、およびモーションキャプチャでの計測を Fig. 4.14 に示す。Fig. 4.14, (a), (b) を見ると、左右の視点で OpenPose の推定が大きく異なっていることが分かる。このデータに関して、右視点が正常に推定されている一方で、左視点では手のリンク構造のうち手首の位置が実際の位置から大きく離れ、人差し指から薬指にかけて指先までのリンクが描画されていない。

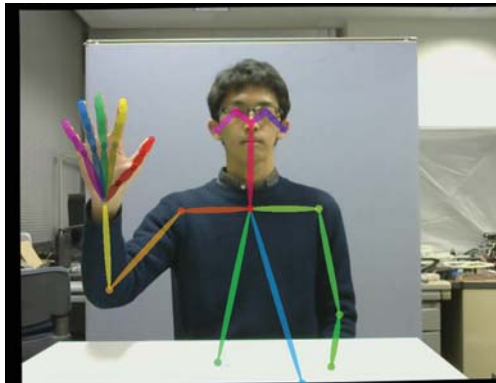
ここで、描画されている関節とそうでない関節について OpenPose の推定結果から尤度を確認すると、前者の尤度が 0.6~0.8 程度であるのに対し、後者では 0.03 や 0.01 といった非常に小さい値であった。本研究では Nelder-Mead 法により OpenPose の推定値を改善しているが、この手法では不十分であることがわかった。したがって、本研究の計測システムでは別の手法による推定値の改善が要求される。

4.4.2 定常誤差に対する考察

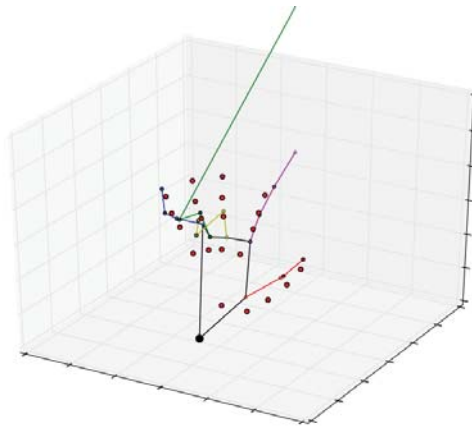
5 指の MCP 関節について、モーションキャプチャでの計測とステレオカメラで復元した 3次元位置の世界座標系における差分を計算する。それぞれの座標系での差分について Fig. 4.15 - 4.17 に、その時の差分平均と標準偏差を Table. 4.9 -



(a) OpenPose estimation (left)



(b) OpenPose estimation (right)



(c) Reconstructed hand and markers

Fig.4.14 Data in the frame with the largest estimation error

4.11 に示す. 各軸の向きは Fig. 2.3 と同様である.

この結果から, この計測システムで推定した手の 3 次元位置は, x, y 軸に関して標準偏差が比較的小さく, 真値であるモーションキャプチャの計測結果に追従できていることが分かる. z 軸に関しては, 各関節での振れ幅が大きく, 奥行き方向の正確な計測が困難であることを示している. しかし, 実験では推定誤差が徐々に収束していることから, 手形状としてはいびつな箇所が見受けられるが, 全体的な手の位置推定を行うことが可能であると考えられる.

4.4.3 手の位置を変更した場合の手形状推定

4.2 節の実験では, 画像の中心に被験者の胴体があり, 手が左側に寄った状態で手形状を推定している. 本項では, 画像上の手の位置によって推定結果にどのような違いが得られるかを確かめるため, 基準となる左視点の画像中心に手を配置して同様の実験を行った. 実験時の手の位置と推定結果について Fig. 4.18 に示す. また, 推定誤差の平均値について全フレームの時系列データを Fig. 4.19 に示す.

実験の結果から, 4.3 節と比較して大きな誤差は無いものの, 同程度の定常誤差が発生していることが分かる. 得られた推定誤差と標準偏差について Table. 4.12 に示す. 他にも, 3.1 節で確認した三角測量の精度を考慮すると, 本章の実験で得られた推定誤差は, OpenPose の推定値に大きく依存することが確認できた.

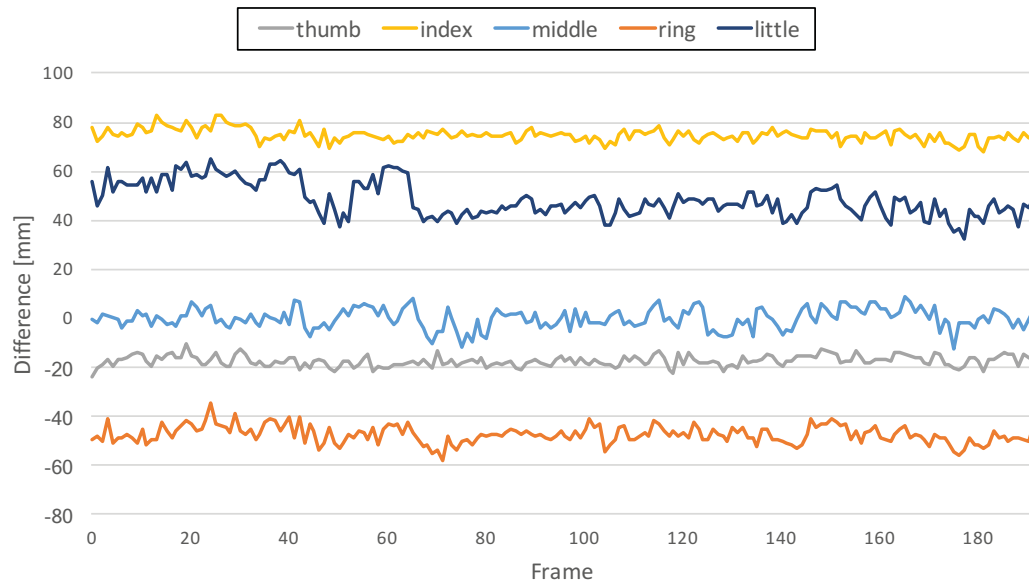


Fig.4.15 Difference in the x axis

Table.4.9 Difference MCP joint position in the x axis

Finger	Mean diferencial [mm]	Standard deviation
thumb	-17.4	2.20
index	74.8	2.53
middle	-0.04	4.10
ring	-47.5	3.56
little	48.5	7.07



Fig.4.16 Difference in the y axis

Table.4.10 Difference MCP joint position in the y axis

Finger	Mean differential [mm]	Standard deviation
thumb	76.4	1.35
index	0.444	1.85
middle	90.5	1.82
ring	47.2	1.85
little	52.6	2.70

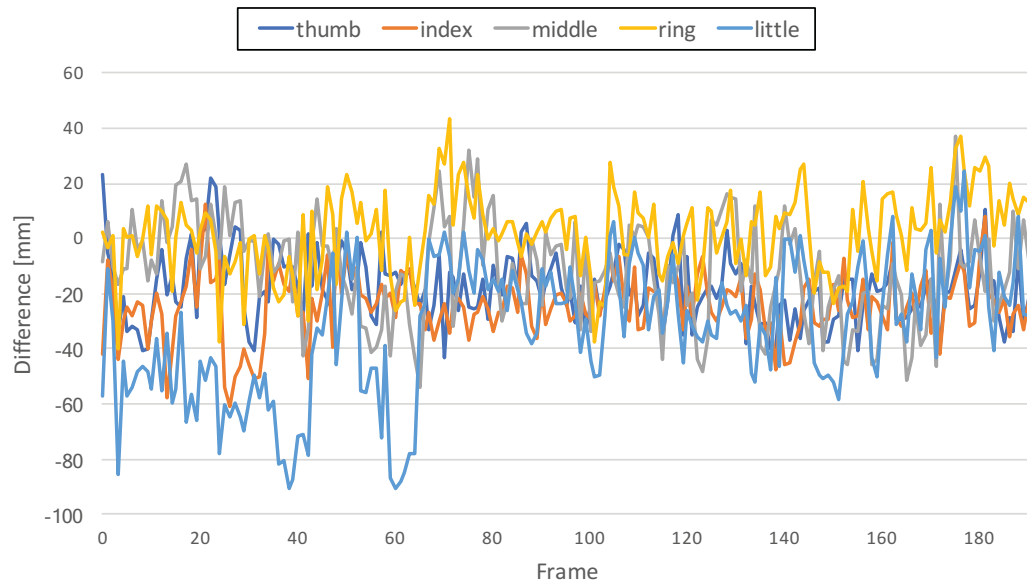
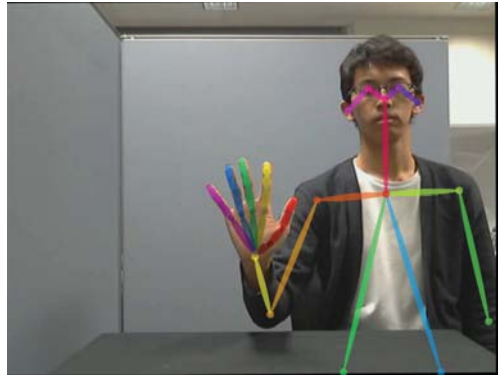


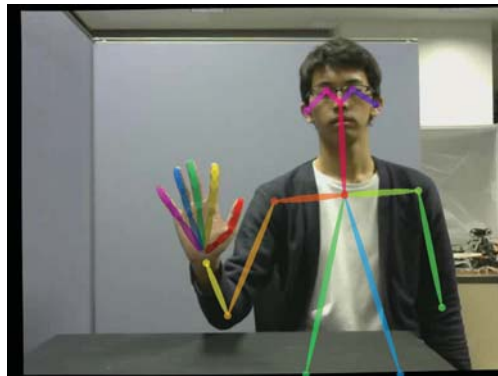
Fig.4.17 Difference in the z axis

Table.4.11 Difference MCP joint position in the z axis

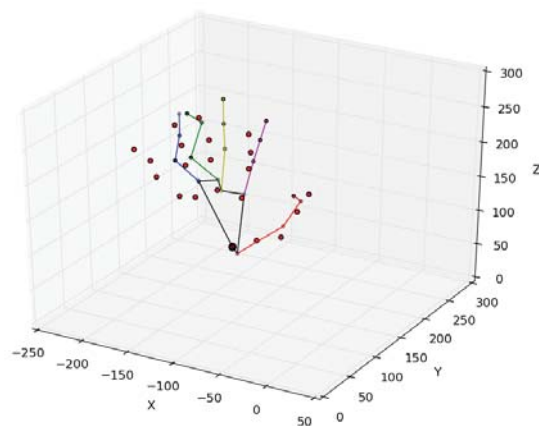
Finger	Mean differential [mm]	Standard deviation
thumb	-18.0	12.5
index	-24.1	11.6
middle	-11.5	18.3
ring	1.89	15.0
little	-32.5	24.2



(a) OpenPose estimation (left)



(b) OpenPose estimation (right)



(c) Reconstructed hand and markers

Fig.4.18 OpenPose estimation from each view and reconstructed hand shape in the case where the hand position is changed

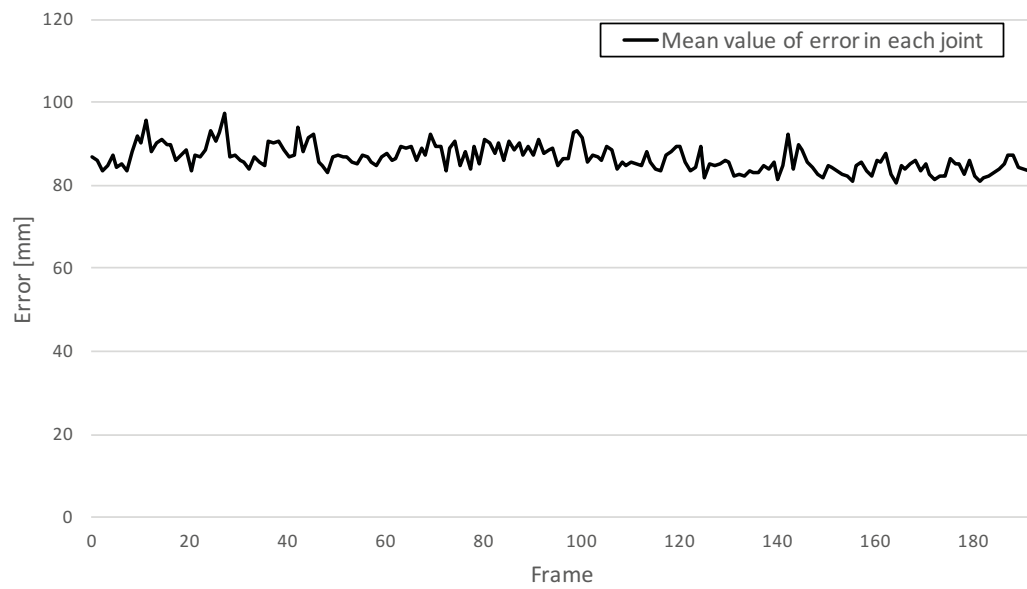


Fig.4.19 Mean value of error in each joint in the case where the hand position is changed

Table.4.12 Estimation error in the case where the hand position is changed

Mean error [mm]	Standard deviation
86.5	3.12

第 5 章 おわりに

本研究では、RGB カメラによるステレオ視の計測システムを構築し、手形状の 3 次元復元を行なった。OpenPose を用いて単眼の RGB 画像から手の 2 次元リンク構造を取得し、ステレオカメラの左右の視点で得られたリンク構造の特徴点を三角測量法によって 3 次元空間に復元した。実験では手をカメラに向かって開き、静止させた状態を計測した。真値としてモーションキャプチャとの計測結果を比較すると、本計測システムでは 80 [mm] 程度の定常誤差は残るが、手形状全体としては実際の手の位置に追従できていることが確認された。

定常誤差に対する考察では、MCP 関節のユークリッド空間における各軸の真値との差と標準偏差を求め、 x, y 軸において差は大きいが標準偏差が小さいことから、これらの軸方向の差が定常誤差に影響していることが確認できた。一方で、 z 軸においては標準偏差が大きく、3 次元に復元した時の手のいびつな形状に影響していることが推察される。しかし、全関節の推定誤差の平均を見ると、定常誤差以外の誤差に大きな影響がないと考えられる。また、手の位置を変えて同様の実験をした結果、上記の実験と同程度の定常誤差が発生した。加えて、本計測システムで OpenPose を用いずに 3 次元位置推定したときの精度と比較することで、OpenPose の推定値が定常誤差の要因であることが推察される。

本研究では、簡単な RGB カメラを用いて計測環境を構築したため、高価なカメラやステレオカメラとしてキャリブレーション済みのデバイスと異なり、様々な誤差が発生した。対策としては OpenPose の推定値を改善することで精度が向上すると推察されるが、そのためにはハードウェアとソフトウェアの両方に余地があると考えられる。ハードに関して、本計測システムでは基線長を 100 [mm] と定めたが、この距離を短くすることで左右の視差を低減し、OpenPose の推定値の差を改善することが考えられる。他にも、背景を白色で統一する、カメラの背後に光源を設置するといったことで、画像中における人の領域とその他の領域を分けることも重要だと考えられる。また、ソフトに関しては、手法として手形状の推定誤差を軽減するために Nelder-Mead 法を実装したが、効果が不十分であった。そのため、推定値の改善のためには別の手法を検討しなければならない。

最後に，実験では手をカメラの正面に向けて静止させた状態で行ったため，手を傾けた状態や動作中において本計測システムで手形状を復元できるかどうかを調査する必要がある．今後は，他の手動作での手形状復元と，上記で示した推定値の誤差に関して実装と実験を通して検証していきたい．

謝辞

本研究は、奈良先端科学技術大学院大学情報科学研究科情報科学先行ロボティクス研究室小笠原司教授のご指導の下で行いました。本研究の遂行にあたり、貴重なご指導とご助言を賜りました小笠原司教授に深く感謝致します。

論文執筆にあたり、ご指導とご助言を頂きました本研究科の杉本謙二教授に深く感謝致します。

また、研究会やミーティングなどで研究方針についてご意見、ご指導頂きました本研究科の高松淳准教授に深く感謝致します。

本研究の研究方針や研究の遂行について多くのご指導、ご助言を頂きました本研究科の丁明助教に深く感謝致します。

大学生活や事務処理など様々な面でサポートして頂きました本研究室秘書の大脇美千代氏、池田美輝氏に深く感謝致します。

本研究室博士後期課程の山崎亘氏には研究や研究生活について多くのご助言を頂きました。深く感謝致します。

本研究室の先輩、同輩、後輩の皆様には研究や学生生活を送るうえで様々なご協力を頂きました。深く感謝致します。

最後に、学生生活を支えて頂いた家族、友人に心から感謝致します。

参考文献

- [1] 上羽康夫： "手その機能と解剖 改定 5 版", 金芳堂, 2010.
- [2] 鎌倉矩子： "手のかたち 手のうごき", 医歯薬出版, 1989.
- [3] J. R. Napier, "The prehensile movements of the human hand," *The Bone and Joint Journal*, vol. 38B, no. 4, pp. 902–913, 1956.
- [4] M.R. Cutkosky, "On Grasp Choice, Grasp Models, and the Design of Hands for Manufacturing Tasks," in *IEEE Transactions on Robotics and Automation*, Vol. 5, no. 3, pp. 269–279, 1989.
- [5] H. Kousaka, H. Mizoguchi, M. Yoshikawa, H. Tanaka and Y. Matsumoto, "Role Analysis of Dominant and Non-dominant Hand in Daily Life," in *IEEE Intl. Conf. on Systems, Man, and Cybernetics (SMC2013)*, pp. 3972–3977, 2013.
- [6] G. Garcia-Hernando, S. Yuan, S. Baek and T. Kim, "First-Person Hand Action Benchmark with RGB-D Videos and 3D HandPoseAnnotations," arXiv, 2017.
- [7] Z. Ju and H. Liu, "Human Hand Motion Analysis With Multi- sensory Information," in *IEEE/ASME Transactions on Mechatronics*, vol. 19, no. 2, pp. 456–466, 2014.
- [8] W. Yamazaki, M. Ding, J. Takamatsu and T. Ogasawara, "Hand Pose Estimation and Motion Recognition Using Egocentric RGB-D Video," in *Proc. of the 2017 IEEE Intl. Conf. on Robotics and Biomimetics (ROBIO 2017)*, pp. 147–152, Macau SAR, China, 2017.

- [9] 伊藤 淳, 吉川 雅博, 丁 明, 高松 淳, 松本 吉央, 小笠原 司, "日常生活における手の動作を記録するウェアラブルシステムの開発", 第 17 回計測自動制御学会システムインテグレーション部門講演会 (SI2016), 2O2-6, 2016.
- [10] H. Joo, T.Simon, X. Li, H. Liu, L. Tan, L. Gui, S. Benerjee, T. Godisart, B. Nabbe, I. Matthews, T. Kanade, S. Nobuhara, and Y. Sheikh, "Panoptic Studio: A Massively Multiview System for Social Interaction Capture," in *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, pp. 1-15, 2017.
- [11] T. Simon, H. Joo, I. Matthews, and Y. Sheikh, "Hand Keypoint Detection in Single Images using Multiview Bootstrapping," in *Proc. of the 2017 IEEE Intl. Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp.1145-1153, 2017.
- [12] CMU-Perceptual-Computing-Lab, "OpenPose," <https://github.com/CMU-Perceptual-Computing-Lab/openpose>, Jan. 2018 Accessed.
- [13] Z. Zhang, "A Flexible New Technique for Camera Calibration," in *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 22, No.11, pp.1330–1334, 2000.
- [14] J. A. Nelder and R. Mead, "A Simplex Method for Function Minimization," *The Computer Journal*, vol. 7, Issue. 4, pp. 308-313, 1965.
- [15] logicool, "HD WEBCAM C270," <https://www.logicool.co.jp/ja-jp/product/hd-webcam-c270>, Jan. 2018 Accessed.
- [16] M. Quigley, K. Conley, B. P. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "Ros: an open-source robot operating system," in

Proc. of the 2009 IEEE Intl. Conf. on Robotics and Automation (ICRA) Workshop on Open Source Software, 2009.

- [17] OptiTrack, "V120 Trio Duo," <https://www.optitrack.co.jp/products/camera/duo.html>, Jan. 2018 Accessed.
- [18] OptiTrack, "Motive," <http://www.mocap.jp/optitrack/products/motive/tracker/index.html>, Jan. 2018 Accessed.
- [19] P. Heo, G. Min Gu, S. Lee, K. Rhee, and J. Kim, "Current Hand Exoskeleton Technologies for Rehabilitation and Assistive Engineering," in *Intl. Journal of Precision Engineering and Manufacturing (IJPEM)*, vol. 13, no. 5, pp 807-824, 2012.