

手指リハビリテーションのための拇指先端可動域の 非接触自動計測システム

浅野 奈生^{†1,a)} 小川 陽子^{†1,b)} 松尾 直志^{†1,c)} 島田 伸敬^{†1,d)}

概要：拇指は他の指と対向することのできる唯一の指であり、外傷時のリハビリテーションにおいては回復度を見積もる上でその指先可動域が重要な指標の一つである。拇指関節の構造は複雑であり指先位置は三次元的に移動するため、従来の道具をあてがって関節の旋回角度を測る方法ではその可動域を正確に計測することが困難である。そこで拇指のリハビリテーションを行う際、拇指先端の三次元可動域を非接触かつ自動的に測定し可視化するシステムを提案する。深度センサーで取得した手全体の深度画像から指先位置を粗く求める CNN モデルと、粗く推定された指先の周辺画素の情報を使って指先位置を正確に求める CNN モデルによって、coarse-to-fine によって親指先位置を推定する。各 CNN の訓練データとして手全体を映した深度画像と指先のカメラ座標系での 3 次元位置の組を与える。推定した指先位置の時系列データをもとに拇指の可動域を三次元的に可視化する。

Contactless measurement system of a thumb tip movement for rehabilitation

NAO ASANO^{†1,a)} YOKO OGAWA^{†1,b)} TADASHI MATSUO^{†1,c)} NOBUTAKA SHIMADA^{†1,d)}

1. 緒言

1.1 研究背景

拇指が負傷した際、外科的治療の後に日常生活に必要な機能を回復するためのリハビリテーションが必要となる。拇指のリハビリテーションを行うにあたり、施術方針を考える上で回復度を把握することは重要である。そのためには現在の手指の状態を知ることが必須である。

手指の回復度を把握する指針としては筋力、関節可動域などがあげられる。この内、拇指の関節可動域では各関節に器具を直接押し当てて関節角度を計測する方法が取られているが、療法士は 1 日に多数の患者の可動域を計測する必要があるため現場では負担となっている。

また、拇指の関節は他の四指よりも複雑で三次元的に可

動するため、従来の道具を押し当てる方法では正確に計測することが困難である。さらに、道具を直接押し当てる計測方法では患者の負担となる可能性もある。

1.2 研究目的

1.1 章で述べた背景を受け、拇指のリハビリテーションを行う際に拇指先端の三次元可動域を非接触かつ自動的に測定し可視化するシステムを提案する。非接触な拇指先端位置計測の実現を目的として、センサから得た情報から機械学習を用いて拇指先端位置を推定できるようにする。また推定で得られたシーケンシャルな拇指先端位置を三次元化・可視化し、複雑な拇指の可動範囲を定量化する。定量化によって、既存の計測方法よりも正確に拇指の回復状況の評価が行える可能性がある。

非接触かつ自動的な計測システムにより、リハビリテーションにおける計測・評価手法の改善をねらう。

^{†1} 現在、立命館大学

Presently with Ritsumeikan University

a) asano@i.ci.ritsumei.ac.jp

b) ogawa@i.ci.ritsumei.ac.jp

c) matsuo@i.ci.ritsumei.ac.jp

d) shimada@i.ci.ritsumei.ac.jp

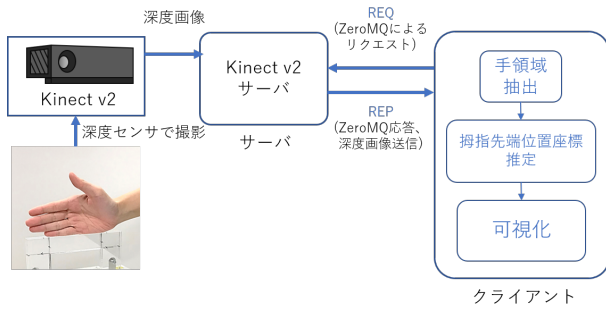


図 1 拇指可動範囲可視化システムの概要

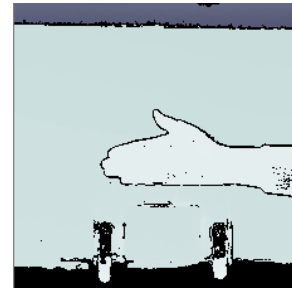


図 2 Kinect v2 で撮影した未加工の深度画像 (一部)



図 3 手領域の抽出を行った深度画像

2. 構築するシステムの概要

システムの概要及びソフトウェアの構成図を図 1 に示す。深度センサとして Microsoft 社の Kinect for Windows v2 (以下 Kinect v2) [4] を使用する。

このシステムでは深度画像から拇指先端座標を推定するプロセスをクライアント、深度センサで撮影した深度画像を推定プロセスに送信するサーバとみなし、同一 PC 内でプロセス間通信を行っている。Kinect v2 から送信される深度画像の 1 画素には 16bit でミリメートル単位の深度情報が格納されている。システム内で使用する CNN モデルの実装・学習には機械学習ライブラリである Chainer[5] を使用した。

また、親指先推定クライアントにより得られた連続する拇指先端座標の推定位置を保存し、これを元に拇指の関節可動域を点群化した手に重ね合わせてリアルタイムに描画することを目的としたモジュールを作成する。

連続した推定座標を元に親指の可動範囲を描画し、手の点群に重ね合わせて描画する。今回は手を Kinect v2 から一定距離に固定して撮影しているため、手の位置のキャリブレーションなどは不要である。

3. 手領域の抽出

3.1 撮影条件

Kinect v2 及びアクリルスタンドを床から 70[cm] の台に配置した。手と Kinect v2 の距離は 60cm とし、深度画像の中心に手が映るようにアクリル製の自作スタンドを作成して固定の上撮影を行った。また、背景オブジェクトによるノイズを撮影時にできるだけ抑えられるよう、1[m] 四方のストレッチボードを Kinect v2 から 80[cm] の距離に置いて撮影した。

3.2 手領域の抽出

Kinect v2 で撮影した未加工の深度画像が図 2 である。未加工の深度画像には手の他にスタンドや背景オブジェクトなどが写り込んでいる。また TOF 方式で測定しているため、手の輪郭部分などで光パルスが拡散し、実際の手

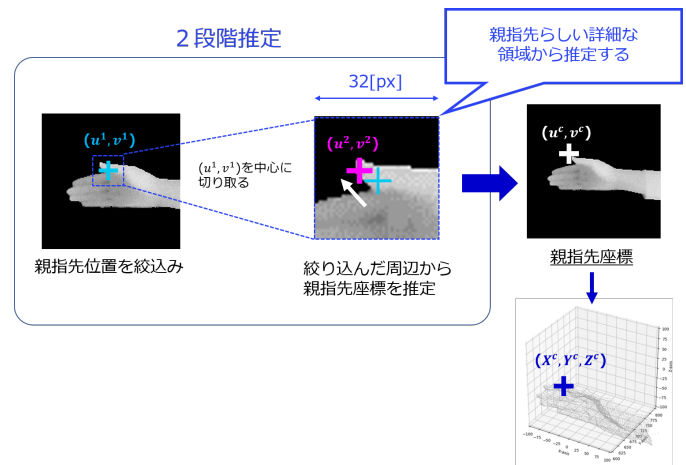


図 4 2段階推定プロセス

よりも大きな距離が計測されてしまうことがある。そして Kinect v2 の性能上、物体内に欠けが生じる事がある。そこで、撮影した深度画像に対しマスク処理による手領域の抽出と深度値の欠損値補完、及び収縮処理による拡散ノイズの除去を行う。

4. 画像平面上の拇指先端位置座標に対する coarse-to-fine な推定

図 4 に 2 つの CNN モデルを用いた coarse-to-fine な推定のプロセスを示す。

拇指先端位置の推定には畳み込みニューラルネットワーク (以下 CNN) を用いる。CNN を用いると、画像内の情報から特徴的な部分を自動的に抽出して学習モデルを構築するため、手の形状や大きさが多少変化しても拇指先端位置を推定することが可能になる。

まず 1 段目の CNN モデルで手が映された深度画像全体の情報を使って拇指先端位置を粗く絞り込んでから、2 段目の CNN モデルで図 4 のように拇指先端周辺の部分深度画像を使って詳細に拇指先端を推定する 2 段階の推定を行う。

また、最終的な推定座標から拇指位置の深度値を得て、これらを 3 次元空間上に投影することで可動域の可視化を行う。

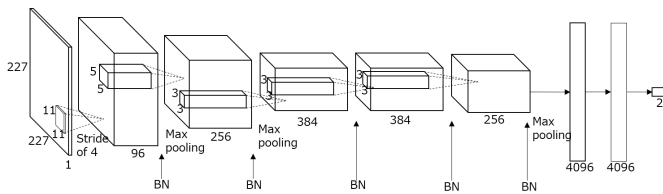


図 5 AlexNet をベースにした CNN(1 段階モデル)

4.1 手全体からの親指先位置絞り込み (1 段階モデル)

1 段階モデルは手全体を映した深度画像から、拇指先端位置座標 (2 次元) を推定する CNN 回帰モデルである。

手全体からの親指先位置を絞り込むことは、深度画像を入力とした回帰問題に相当する。そこで、AlexNet[6] を、深度画像から拇指先端位置座標 (u, v) の回帰に用いる。今研究では 2 次元座標を推定することを目的に出力を 2 ノードに変更した。

また、CNN への入力を 3 チャンネルから 1 チャンネルに変更し、全ての畳み込み層の後に Batch Normalization[7] を追加した。以上の変更を反映した、1 段階モデルの構成を図 5 に示す。

本研究では 1 人の手に対して 182 枚の深度画像を撮影し、 227×227 [px]、4550 枚の学習用画像を作成した。また作成した深度画像に対応する教師信号 (u, v) を作成する。教師信号の位置は目視によって与えた。学習用画像の内 $\frac{2}{3}$ を学習セット、 $\frac{1}{3}$ をテストセットとし、学習用画像を入力として推定座標を出力するよう、CNN の学習を行った。

4.2 親指先周辺画像からの親指先位置推定 (2 段階モデル)

2 段階モデルは親指先周辺で切り取った部分深度画像から、拇指先端位置座標 (2 次元) を推定する CNN モデルである。

2 段階モデルの構成を図 6 に示す。手全体からの親指先位置を絞り込むことは、深度画像を入力とした 2 クラス分類問題に相当する。このモデルを、深度画像から拇指先端位置座標 (u, v) の回帰に用いる。

1 人の手を撮影した深度画像 1678 枚から、 32×32 [px]、41950 枚の学習用画像を作成した。また作成した部分深度画像に対応する相対的な拇指先端位置座標を求め、教師信号とした。学習用画像の内 $\frac{2}{3}$ を学習セット、 $\frac{1}{3}$ をテストセットとし、拇指先端位置座標を出力するよう、CNN の学習を行った。

4.3 2 つの CNN モデルを用いる 2 段階推定の性能検証

1 段階モデルの学習に用いたものと同じデータセットを使用する。このデータセットの内 $\frac{2}{3}$ を学習セット、 $\frac{1}{3}$ をテストセットとする。各セット内の学習用画像を入力として coarse-to-fine な推定が行えるか検証を行った。

推定の 1 段階には 4.1 節、2 段階には 4.2 節で作成したモ

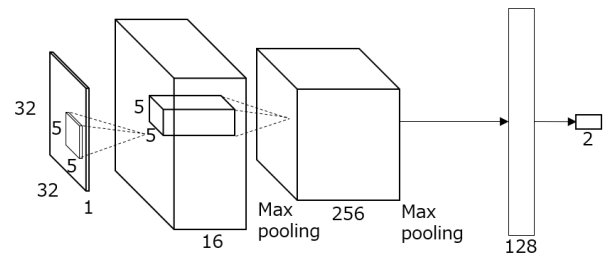


図 6 親指先周辺画像から親指先位置推定に使用するモデル (2 段階モデル)

デルを用いる。 227×227 [px] の深度画像を入力とした、1 段階 CNN モデルの出力 (推定座標) を中心として 32×32 [px] の小領域をトリミングする。トリミングした画像に対する 2 段階 CNN モデルの出力 (推定座標) は元の深度画像に対する相対的な推定座標となる。これを絶対座標に変換し、最終的な推定座標を得る。

4.3.1 推定結果

学習性能と汎化性能の評価基準として、各セットに対する平均二乗誤差を計算した。(表 1) 1 段階モデルのみで推定した場合の平均二乗誤差も合わせて掲載した。平均二乗誤差の平方根を取ると、深度画像上における平均の誤差 [px] が求められる。丸括弧内に平方根を記載している。

深度画像に 2 段階推定による推定結果を重ねた例を図 7 に示す。上 2 枚が学習セット、下 2 枚がテストセットの推定例である。白色の十字が教師信号、水色の十字が 1 段階目の推定モデルによる拇指先端位置の推定座標、ピンク色の十字が 2 段階推定による拇指先端位置の推定座標である。また、本来の画像から手領域周辺をトリミングして掲載した。

表 1 各セットに対する平均二乗誤差

	学習セット	テストセット
2 段階推定	10.52 (3.24)	10.57 (3.25)
1 段階のみ	134.66 (11.60)	133.29 (11.54)

4.4 考察

推定結果の例を見たところ、各セットの全ての画像で拇指先端位置をある程度推定できていることが確認できた。

また、平均二乗誤差の計算結果から 2 段階推定と 1 段階モデルのみでの推定性能を比較した。表 1 を見ると 1 段階モデルのみでは平均して 11[px] の誤差が出るが、2 段階推定では 3[px] 以内に誤差が減少したことから 2 段階推定を行うことで推定性能が上がったことが分かる。

5. 3 次元空間上の拇指先端位置座標に対する coarse-to-fine な推定

章での推定結果を 3 次元空間上に投影したものが図 8 で

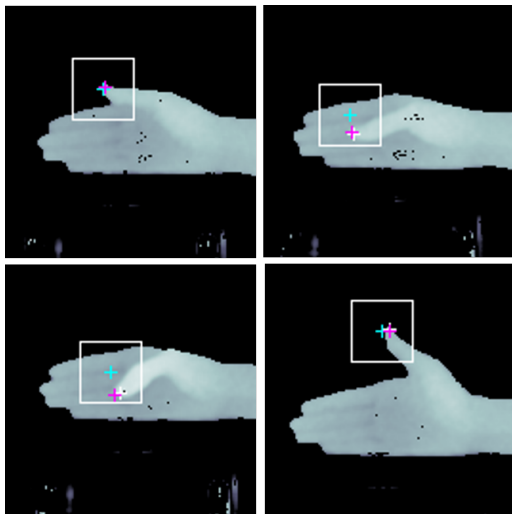


図 7 学習セット、テストセットの 2 段階推定例

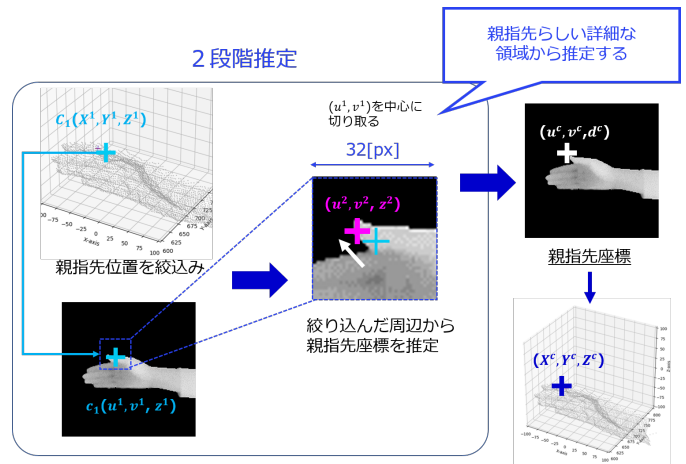


図 9 2 段階推定プロセス

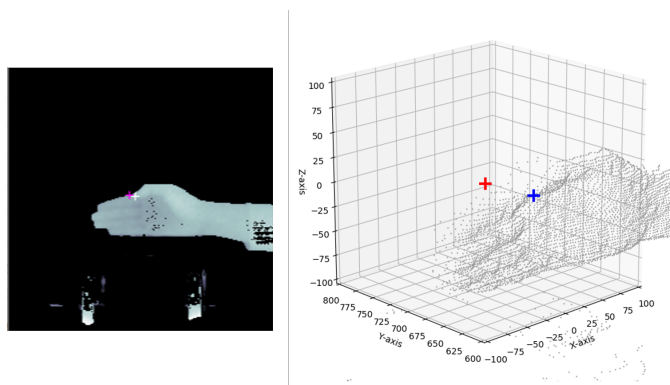


図 8 推定座標に対応する深度値が、実際には手領域と離れた位置にある例

ある。図 8 には 3 次元点群化した手の上に推定座標に対応する点を赤色で描画した。図 8 のように、2 次元平面上では拇指先端座標を推定できているように見えても、推定座標に対応する深度値は手領域とは離れた位置にある場合がある。

そのため、直接に 3 次元空間上の拇指先端位置座標を推定することを考える。深度画像内の画素値に依存しない 3 次元座標を推定できるようにすれば、図 8 で示したようなケースでも真の拇指先端位置を推定することができる。

図 9 に 2 段階推定のプロセスを示す。1 段目で 3 次元空間上の拇指先端周辺位置 $C_1(X_1, Y_1, Z_1)$ を絞り込み、拇指先端周辺の小領域を抽出する。2 段目では小領域に対して CNN モデルによって画像上の拇指先端座標 $c_2(u_2, v_2, D_2)$ を推定する。 C_1 と c_2 の関係から画像上の座標 $c(u, v, D)$ を求め、それを 3 次元空間座標に変換した $C(X, Y, Z)$ を推定結果とする。

5.1 手全体の深度画像から指先位置を回帰する CNN モデル (1 段階モデル)

手領域を抽出した深度画像に対応した、実空間上にお

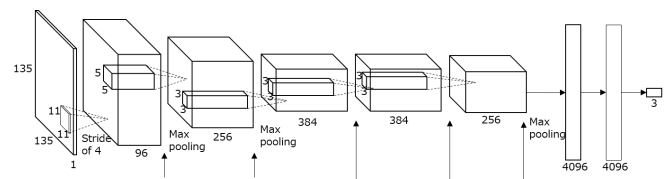


図 10 AlexNet をベースにした CNN

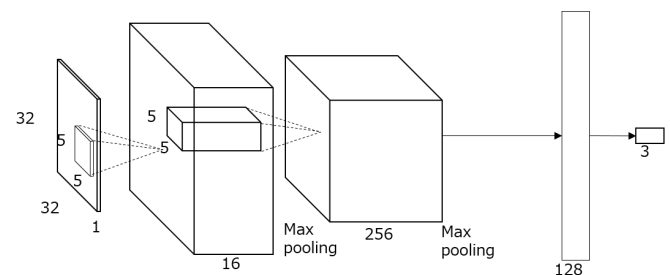


図 11 使用したモデル

る母指先端位置の 3 次元座標 $C_1(X_1, Y_1, Z_1)$ を教師信号とし、この深度画像と教師信号の組からなる学習セットを作成する。この学習セットを用い、深度画像を入力すると母指先端位置の 3 次元座標を推定するように CNN を学習する。

学習には AlexNet[6] をベースにした CNN を使用する。今研究では 3 次元座標を推定することを目的に出力を 3 ノードに変更した。変更を反映した、1 段階 CNN モデルの構成を図 10 に示す。

5.2 指先周辺の深度画素の情報から指先位置を回帰する CNN モデル (2 段階モデル)

画像上の座標 $c_2(u_2, v_2, D_2)$ を推定することを目的に、出力を 3 ノードに変更する。また全層の後に活性化関数を追加する。今研究では ReLU[8] を用いる。以上の変更を反映した、2 段階 CNN モデルの構成を図 11 に示す。

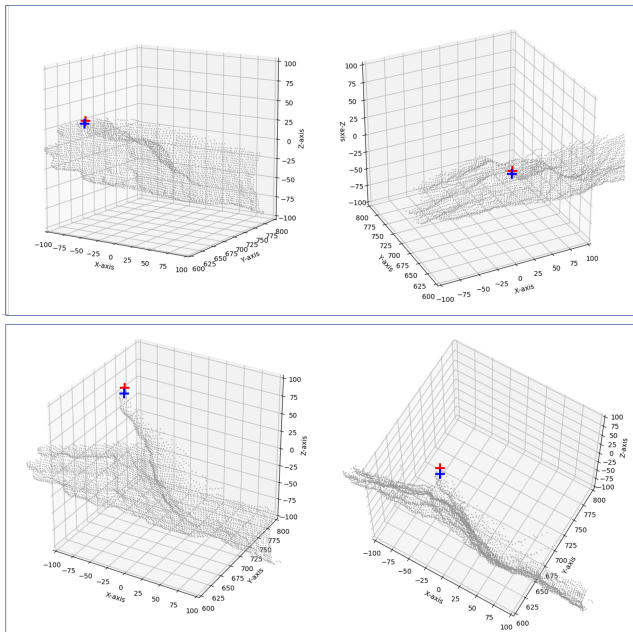


図 12 3次元空間上に点群化した学習セットの深度画像と教師信号、推定座標を描画した例

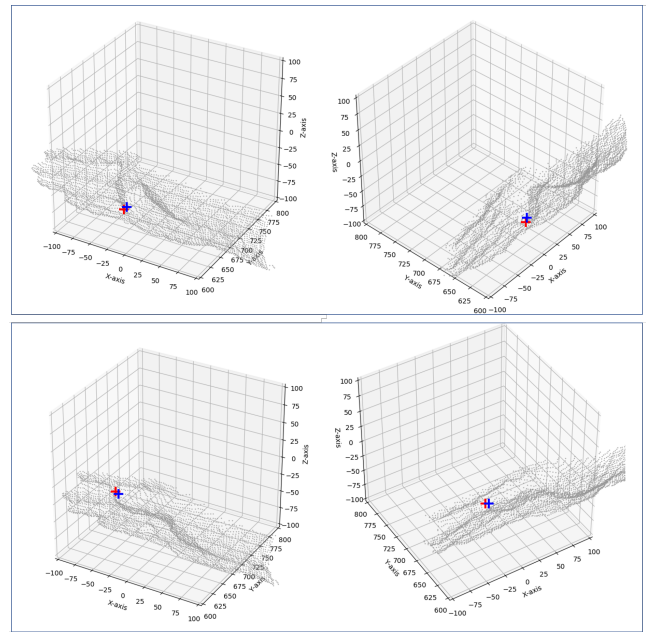


図 13 3次元空間上に点群化したテストセットの深度画像と教師信号、推定座標を描画した例

5.3 2つのCNNモデルを用いる2段階推定の性能検証

1段階CNNモデルの学習に用いたものと同じデータセットを使用する。このデータセットの内 $\frac{2}{3}$ を学習セット、 $\frac{1}{3}$ をテストセットとする。各セット内の学習用画像を入力として coarse-to-fine な推定が行えるか検証を行った。

推定の1段階には5.1節,2段階には5.2節で作成したモデルを用いる。

5.3.1 推定結果

学習性能と汎化性能の評価基準として、各セットに対する平均二乗誤差を計算した。その結果を表2に示す。また、表2には1段階モデルのみで推定した場合の平均二乗誤差も合わせて掲載した。なお、平均二乗誤差の平方根を取ると誤差 [mm] に相当する。

図12、13に3次元空間上における拇指先端位置座標を2段階推定によって推定した例を示す。3次元空間上に正解を赤色の十字、推定座標を青色の十字で描画した。また図12、13について、同じ深度画像を元にした点群を別角度から見たものを枠で囲ってある。

表 2 各セットに対する平均二乗誤差

	学習セット	テストセット
2段階推定	33.55	32.68
1段階のみ	15.42	18.36

5.3.2 考察

3次元座標を推定する2段階推定について、平均二乗誤差の計算結果を見ると1段階モデルのみによる推定結果よりも学習性能、汎化性能共に悪くなっている。

また推定結果の例からは、1段階モデルでよく推定できていた画像に対しても、2段階推定を行うことで誤差が大

きくなってしまうことが分かる。

6. 別人物の手に対する実験

作成したシステムを用いて、作成した学習済みCNNモデルを使用し、提案する2段階推定の有効性を検証する。

前章までにモデルの学習に用いたのは1人の手を撮影した深度画像であった。そこで学習に使った深度画像の被写体と異なる人物の手に対して連続的に深度画像を撮影しても、母指先端位置座標が推定できるかを確認する。

今研究では、別人物の手に対して3次元空間上の拇指先端位置座標を推定した。

6.1 実験環境

6.1.1 深度画像の撮影環境

実験時には掌を床に対し水平に固定できるようにアクリルスタンドの上に被験者の右手を置き撮影を行う。Kinect v2とアクリルスタンドの距離は60[cm]とする。

また、撮影時には深度センサの中央に掌が映るように調整の上、実験を行った。

6.1.2 被験者データ

今実験ではモデルの学習に用いた深度画像の被写体を被験者Xと呼ぶ。また、掌の大きさの異なる被験者A,B,Cの3名の協力を得て行った。表3に被験者の各データを示す。表3に掲載した掌のサイズは手首から中指先の長さである。また、拇指の長さは第一中手骨末端から拇指先端までの長さである。

表 3 被験者データ

	性別	年齢	掌のサイズ [cm]	拇指の長さ [cm]
被験者 X	女	23	17.5	10.0
被験者 A	男	24	17.0	9.5
被験者 B	男	23	18.0	10.0
被験者 C	男	23	18.0	10.0

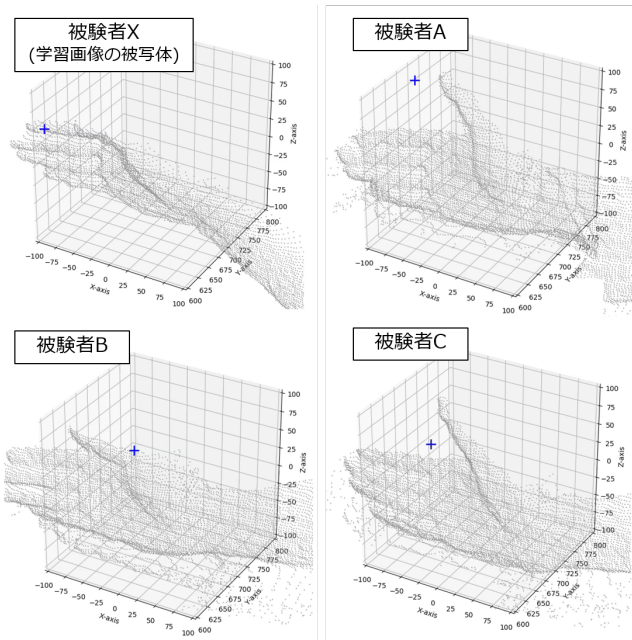


図 14 別人物の手に対する推定結果の例

6.2 実験手順

6.2.1 実験内容の説明

実験は以下の手順で行う。

- (1) 被験者にアクリルスタンドの上に右手を載せ動作を行ってもらう (練習)
- (2) Kinect v2 で深度画像を撮影しながら、再度動作を行ってもらう
- (3) 動画の各フレームに対し 2 段階推定を用いて 3 次元空間上の拇指先端位置座標を求める
- (4) 時系列で得られた拇指先端座標を描画する

6.3 実験結果

6.3.1 推定結果の例

被験者 X,A,B,C の動作に対してリアルタイムに推定し、時系列順に連続した深度画像と拇指先端位置座標を得た。その一部の例が図 14 である。3 次元化した点群 (灰色) の上に、得られた推定座標を青色の点で描画した。

6.4 考察

CNN モデル作成時に撮影した被験者 X の手に対しては、かなり拇指先端に近い位置に推定することもあるが、多くは外れてしまっていた。また、CNN モデル作成時とは異なる形状の手に対しては概ね推定できるものと、ほとん

ど推定できないものがあることが確認できた。

スタンドに掌を固定してもらった状態では、被験者によって掌形状が大きく変形した。このため、拇指とは全く違う場所に推定座標が存在することがあった。十分に手形状の特徴から拇指先端位置座標を推定できなかったためと考えられ、複数人を撮影対象とした画像を 1 段目の CNN モデルに学習させることで全体の精度が上がることを期待できる。

7. 結言

7.1 まとめ

本研究では拇指先端の三次元可動域を非接触かつ自動的に測定し可視化するシステムを提案した。深度センサで撮影した深度画像を用い、拇指の可動範囲を三次元化・可視化する方法について示した。また、実験ではある程度形状の異なる手に対しても、拇指の可動範囲を推定することができるか検証した。

7.2 今後の課題

今後の課題として、CNN モデル作成に関しては学習用画像の追加があげられる。本研究で作成した学習用画像は 1 種類の手を撮影したものであった。実際のリハビリテーションの現場で用いるには様々な形状・病状の手に対応するため、複数の形状の異なる手を撮影し学習に用いる必要がある。また可視化方法に関し、今回は点群及び凸包を用いたが、手の状態の比較・診断を行えるようにするシステム構築のため療法士にとって扱いやすい UI・UX を検討する。

参考文献

- [1] 日本理学療法士協会 学会版 MMT および学会版 ROM 評価指針に関するパブリックコメントの募集 資料 3. 関節可動域測定法 (本編), 2018 年 1 月 28 日閲覧 http://www.japanpt.or.jp/upload/jspt/obj/files/publiccomment/3_rom_20140612.pdf
- [2] 米本恭三, 石神重信, 近藤徹 "関節可動域表示ならびに測定法." リハビリテーション医学, 32.4 (1995): 207-217.
- [3] 石田和宏 "日本理学療法士学会版の関節可動域評価指針." 理学療法学 42.8 (2015): 763-764.
- [4] Developing with Kinect, 2018 年 1 月 22 日閲覧 <https://dev.windows.com/en-us/kinect/develop>
- [5] Tokui, S., Oono, K., Hido, S., & Clayton, J. "Chainer: a next-generation open source framework for deep learning." Proceedings of workshop on machine learning systems (LearningSys) in the twenty-ninth annual conference on neural information processing systems (NIPS). Vol. 5. 2015.
- [6] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks." In NIPS, 2012.
- [7] Ioffe, Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." International Conference on Machine Learning. 2015.

- [8] Xavier Glorot; Antoine Bordes; Yoshua Bengio. “Deep Sparse Rectifier Neural Networks” . Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (AISTATS-11) 15: 315-323.