

リハビリテーションのための親指先端可動域の計測と可視化

3-D Measurement and Visualization of Thumb-tip for Medical Care and Rehabilitations

非 三宅雄大（立命大） 非 松尾直志（立命大） 非○島田伸敬（立命大）
Yuta MIYAKE, Tadashi MATSUO, Nobutaka SHIAMDA (Ritsumeikan University)

For medical care and rehabilitation of hand and finger injury we developed a 3-D measurement and visualization framework and prototype system, which employs RGBD sensor and a deep neural network based visual detection and tracking method. By using this system for the actual medical treatment, doctors and medical staffs can record patient's finger ability in quantitatively and perceive the effectiveness of the treatment intuitively. The experimental results for virtual testers are shown.

Key Words: RGBD sensor, Deep neural network, Computer vision

1 はじめに

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望ましい。手指の回復度を把握する指針として筋力・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。

本稿では拇指のリハビリテーションを行う際に拇指先端の3次元可動域を非接触に測定し可視化するシステムを提案する。非接触な拇指先端座標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座標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座標を3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。

現在のリハビリテーションにおける、関節可動域の計測手法について述べる。標準的な測定法としては、日本整形外科学会と日本リハビリテーション学会による「関節可動域表示並びに測定法(1995)」(以下 H7 測定法)[1] を基本とし、H7 測定法で定められた関節位置にゴニオメーターと呼ばれる分度器状の器具を測定したい関節に当て、関節可動域を測定する。しかし、石田[2]によると多くの療法士は自身の実習・臨床経験に基づいた H7 測定法をベースとした独自の測定法を用いているのが現状である。また実習・臨床経験の違いやゴニオメーターの異なる当て方によって誤差が発生してしまう可能性があると述べられている。このため現在の測定法では、関節可動域の測定値が療法士の技量に依存してしまう。

2 撮影条件及びシステム構成

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。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 の性能を考慮し、事前に手の深度画像の撮影条件を定めた。手と Kinect v2 の距離は 60[cm] とし、深度画像の中心に手が映るように 2 つの箱を固定した高さ 10[cm] の位置で撮影を行った。また、撮影時に手の周りの背景オブジェクトによるノイズを抑えられるよう、冊子を Kinect v2 から 80[cm] の距離に置いて撮影した。撮影時の様子を図 2 に示す。

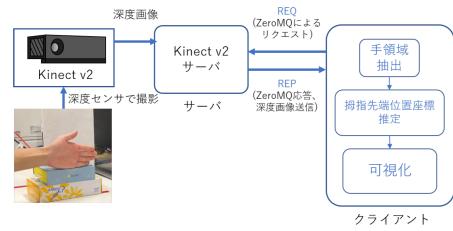


Fig.1 Proposed measurement framework

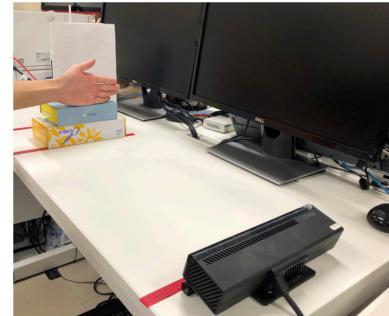


Fig.2 System Prototype

3 手指画像の撮影

3.1 撮影する動作

被験者の手指を撮影する動作は以下の通りである。1 から 4 までの動作 (以下測定動作) を数回行った後、拇指対向動作を各指に 1 回ずつ各被験者に行ってもらった。

1. 横側外転・尺側内転
2. 掌側外転・掌側内転
3. 屈曲・進展 (MP)
4. 屈曲・進展 (IP)
5. 拇指対向動作 (示指、中指、薬指、小指)

3.2 マスク処理

Kinect v2 で撮影した深度画像が図 3 である。本来、深度画像は 16[bit] で深度情報を記録しているが、未加工の状態では目視で何が映っているかを判断することが難しい。そのため、目視で確認する必要がある場合は 8[bit] 化して可視化する。この深度画像には手の他に台や背景オブジェクトなどが写り込んでいるため、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を行う。はじめに台の部分をマスク処理によって除去する。次に、撮影条件及び拇指の可動範囲を考慮し、深度画像内における深度値が 550[mm] から 705[mm] の領域を抽出し、手の背景部分は深度値を 0 で埋める。同時に、撮影した深度画像の中央を基準に手領域のみを抽出した深度画像を生成する。これは、撮影した深度

画像は Kinect v2 の性能で $512 \times 424[\text{px}]$ であるが、機械学習に使用するために $227 \times 227[\text{px}]$ にする必要があるためである。

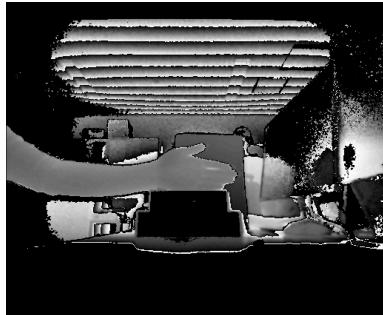


Fig.3 Depth image

3.3 訓練用画像セットの準備

前節で撮影した画像に対して拇指の X, Y 座標を機械学習によって推定するには、教師信号として正しい拇指の座標を与える必要がある。機械学習に使用する画像（以降学習画像）はすべて $16[\text{bit}]$ 深度の画像を使用しているが、ディスプレイ上で目視しやすいように $24[\text{bit}]$ 深度に変換し 1 枚ずつ画像を表示し拇指の位置をクリックしていくことで指定する。

機械学習には多数の画像を用いることで、物体の特徴を学習し性能を上げることができる。そのため本研究では効率的に多数の画像を用意するため、1 枚の深度画像から複数枚の深度画像を切り出してデータセットを作成する。図 4 のように、1 枚の深度画像から画像中央を基準として縦横に $4[\text{px}]$ づつずらし、合計 25 枚の $227 \times 227[\text{px}]$ の学習画像を作成する。そして得られた学習画像から時系列順に、1 枚を検証データ（validation に使用するデータ）、2 枚を学習データ（予測モデルの作成に使用するデータ）と交互に分ける。

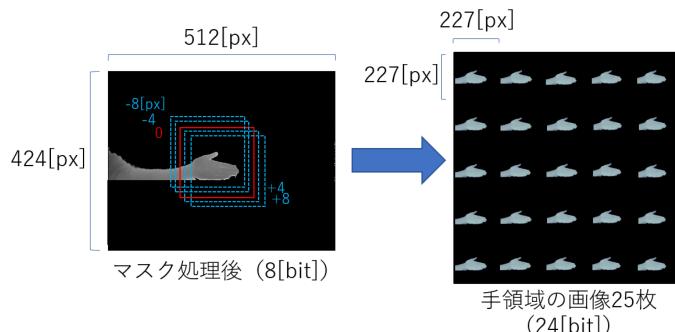


Fig.4 Data segmentation

被験者 A,B,C,D,E,F,G の 7 名の協力を得て学習及び検証用のデータ収集を行った。表 1 に被験者の各データを示す。なお、被験者全員に発表論文に各データ及び実験データを匿名で掲載することの同意を得てからデータを取得した。表 1 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。図 5 に拇指の各骨と関節位置を示す。今回の実験では、表表 1 のうち被験者 A,B,C,D,E,F のデータを 2 つに分け学習画像とテスト画像とし、被験者 G のデータは全てテスト画像とした。結果、学習画像である被験者 A,B,C,D,E,F,G のデータは計 15250 枚であり、そのうち学習用データが 10165 枚、検証用データが 5085 枚である。うち、被験者 A のデータは 2600 枚、被験者 B は 2450 枚、被験者 C は 2675 枚、被験者 D は 2750 枚、被験者 E は 2250 枚、被験者 F は 2875 枚、被験者 G は 2325 枚である（図 6 に一部を表示）。

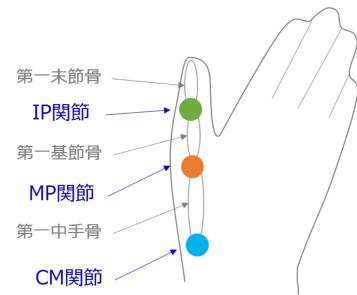


Fig.5 Finger joints

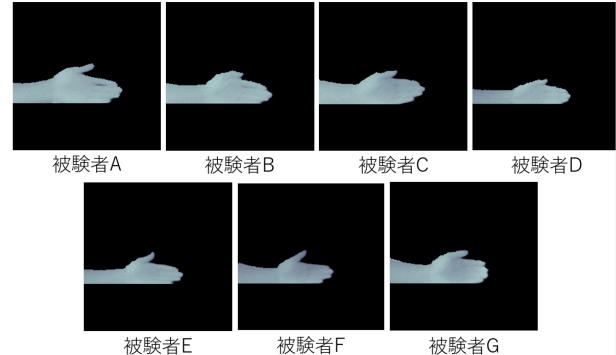


Fig.6 Training images

4 DNN による親指先位置の検出

4.1 ResNet-50

ResNet (Residual Network) [3] は CNN 同様画像の分類や回帰を行うニューラルネットワークである。層を深くすると勾配消失問題、劣化問題が発生して学習が難しくなるが、深い層でも学習が行えるネットワークアーキテクチャとして入力をそのまま出力にバイパスして合流させる Res Block を導入したものが ResNet である。ResNet は、ResNet-18 (18 層)、ResNet-34 (34 層)、ResNet-50 (50 層)、ResNet-101 (101 層)、ResNet-152 (152 層) の 5 種類が提案されており、本研究では ResNet-50 を使用する。

ResNet の訓練に用いる損失関数は平均二乗誤差を用いた。

4.2 訓練

過学習や学習不足を防ぐため、更新エポックごとのモデルの損失関数の値を確かめた。今回用いた学習データ、検証データの損失関数の値の変化の様子を図 7 に示す。縦軸が損失、横軸がエポック数を表している。

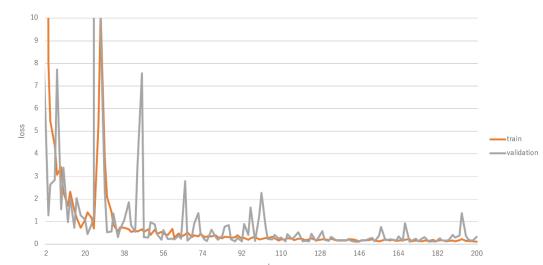


Fig.7 Convergence of loss values

被験者 7 名のデータに対する損失関数の値（画素位置の平均二乗誤差）を表 2 に示す。訓練を使った被験者（ただし画像 자체は訓練データに含めなかったもの）である A,B,C,D,E,F の平均二乗誤差は非常に小さく、 $1[\text{px}]$ 未満であり教師信号と推定座標がほぼ一致している。訓練には一切に用いなかった被験者 G のデータについても約 $3[\text{px}]$ にとどまっている。

Table 1 Finger spec. of each subject

	性別	掌のサイズ [cm]	拇指の長さ [cm]
被験者 A	男	19.0	10.5
被験者 B	男	18.0	10.0
被験者 C	男	19.5	11.0
被験者 D	女	16.0	9.5
被験者 E	女	17.0	9.5
被験者 F	女	16.5	9.0
被験者 G	男	18.0	9.5

Table 2 Estimation error for each subject

	平均二乗誤差 [px ²]	平方根 [px]
被験者 A	0.18	0.42
被験者 B	0.22	0.46
被験者 C	0.18	0.42
被験者 D	0.28	0.52
被験者 E	0.11	0.33
被験者 F	0.17	0.41
被験者 G	9.23	3.04

被験者 7 名の訓練データに対する誤差ヒストグラムを図 8 に示す。これらのグラフの縦軸は、全画像に対する誤差の割合を表しており、横軸は、例えば 0~1 の場合、誤差が 0[px] 以上 1[px] 未満といったように誤差の範囲を表している。

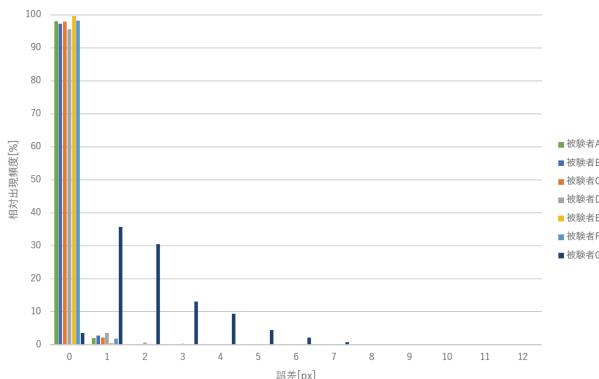


Fig.8 Error histogram

4.3 親指先位置の推定結果と 3 次元可視化

平均二乗誤差、誤差の相対出現頻度の結果を踏まえ、被験者 A,B,C,D,E,F をテスト画像とした場合、学習に使用した画像では結果に大きな差が生じないと判断した。そのため、被験者 A,B,C,D,E,F のうち被験者 A を代表として、被験者 A,G の推定結果の例を図 9、図 10 に示す。図中の白色の十字は教師信号、ピンク色の十字は推定座標を表しており、各図の左上の数値は推定座標 [x, y] を表している。これらの図は、特に手や拇指の形状が異なる場合を選択している。

被験者 A と G の推定結果を 3 次元空間上で可視化したものを見ると、図 11、図 12 に示す。各図は、それぞれ図 9、図 10 を 3 次元空間上で可視化したものである。赤い十字は教師信号、青い点は推定座標を表している。

図 8 に示す誤差ヒストグラムを見ると、被験者 A,B,C,D,E,F のデータでは 0[px] 以上 1[px] 未満の誤差が 90[%] 以上となっており、ほとんど誤差がないことがわかる。被験者 G のデータでは 0[px] 以上 1[px] 未満の誤差が 5[%] にも満たないものの、1[px] 以上 3[px] 未満の誤差が約 80[%] を占める結果となっている。図 9、図 10 より、被験者 A のデータでは、手や拇指がどの形状にあっても誤差は微小である。被験者 G のデータでは、訓練データに含まれている被験者 A より大きいもののやはり誤差は小さい。

図 11、図 12 より、X,Y 座標に関しては、3 次元座標系でも図 9、図 10 に示す画像座標系と同じ位置を再現できている。一

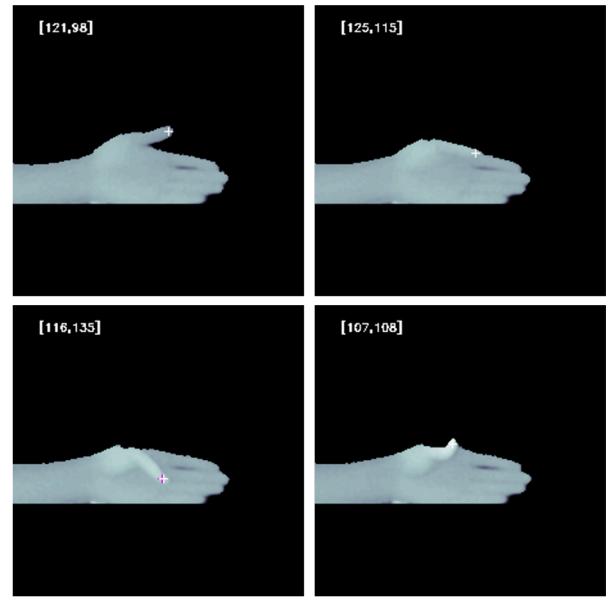


Fig.9 Estimation results for Subject A (included in training)

方、奥行きの Z 座標に関しては、X,Y 座標よりも大きめの誤差が出ている。

以上の検証より、訓練に使用した被験者の同じ拇指の動作に対しては極めて正確な精度が達成されており、学習画像には含まれない未知の手や拇指の形状を持つ人に対しても良好な結果が得られることがわかった。さらに多くの被験者について拇指のデータを訓練に使用し、拇指の動作の種類も増やせば、未知のデータに対してより精度が上がる事が期待される。

また、リハビリ開始当初とリハビリ後の比較を行うことを想定して、被験者に親指で小さな動きと大きな動きの両方をしてもらったものを 3 次元可視化し、重ねて表示したものを図 13 に示す。3 次元的に表示ができる直感的にその稼働範囲の差が分かりやすく示されていることがわかる。実際のアプリケーションではマウス等で表示の視点位置をリアルタイムに移動させることも可能である。

5 まとめ

本研究では、深度センサで撮影した深度画像を用いて機械学習を中心に、3 次元空間上で拇指先端座標を計測し可視化することを目的とした。機械学習による拇指先端座標の計測、3 次元空間上での可視化において、先行研究の方法と比較し検証した。実験では複数の種類の学習画像を用いて、形状の異なる手または拇指に対しても拇指先端座標を推定することができるか検証した。また、実際の現場で使うことを想定し、拇指先端可動域を描画する方法について提案した。

本研究で作成した学習画像は 6 人の被験者の同じ動作を撮影したものであったが、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、より多くの被験者と様々な拇指の動作を学習画像に含む必要がある。また、3 次元空間上での可視化方法に関しては、今回は 1 つの 3 次元空間上に色分けした 2 つのデータの推定座標を点で描画ただけであったが、現場の療法士にとってより扱いやすい UI・UX を検討する。

参考文献

- [1] 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, pp.207-217, 1995.
- [2] 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, pp.763-764, 2015.
- [3] K. HE, et al. "Deep residual learning for image recognition", Proc. of IEEE conf. on computer vision and pattern recognition, pp.770-778, 2016.

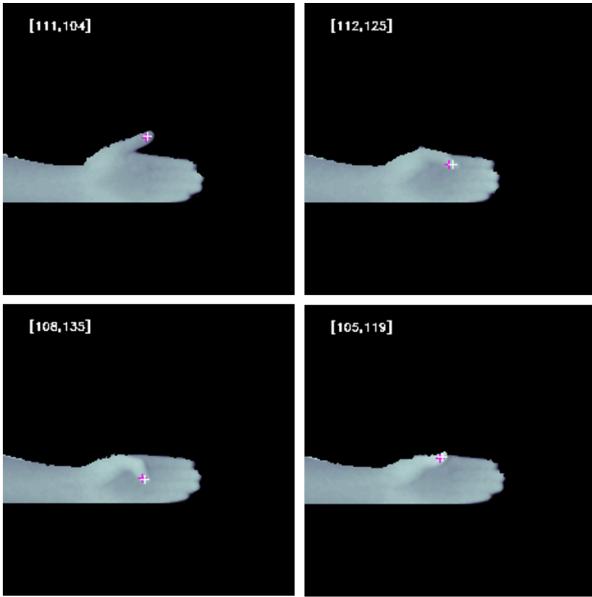


Fig.10 Estimation results for Subject G (not included in trainging)

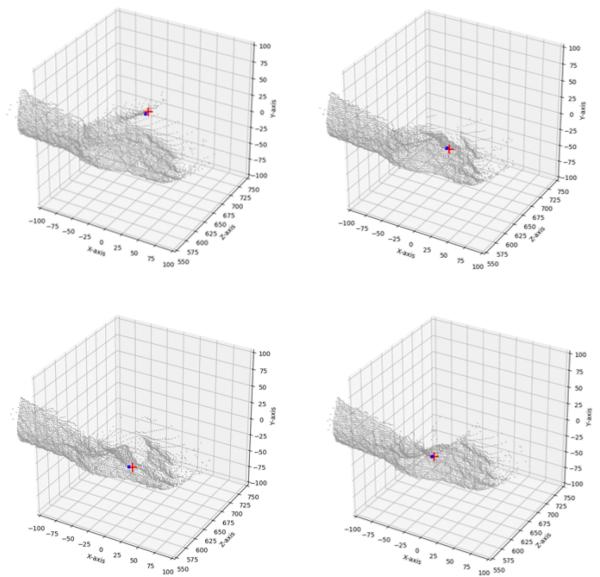


Fig.12 3-D view for Subject G (not included in trainging)

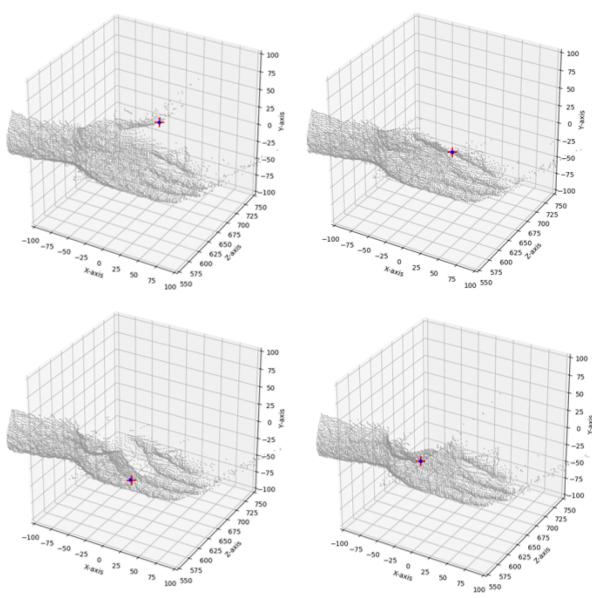


Fig.11 3-D view for Subject A (included in trainging)

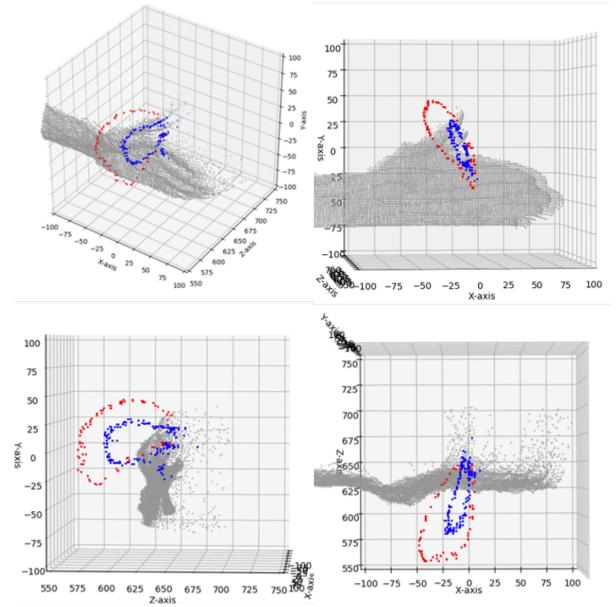


Fig.13 Comparison between two different tracks