



磁気式モーションセンサと CNN を用いたオクルージョンに強い非グローブ型手形状・位置・姿勢推定手法

菅原 諒¹⁾, 黄 佳維¹⁾, 高嶋 和毅¹⁾, 北村 喜文¹⁾

1) 東北大学 電気通信研究所 (〒980-8577 宮城県仙台市青葉区片平 2-1-1, icd-imxd@grp.tohoku.ac.jp)

概要: 限られた数の磁気式モーションセンサの出力を機械学習し, 細かい手指の動作を行う手の代表点の位置・姿勢および各指の関節間の全部位の姿勢を推定する方法を提案する. 磁気式モーションセンサにより, オクルージョンが生じる状況やグローブが使用できない状況でも対象の推定が可能となる. 機械学習のモデルには, 2次元格子状に配置された磁気式モーションセンサの計測値を入力とする CNN のモデルを採用した. 磁気式モーションセンサの 7 つのマーカを片手に装着し, OptiTrack と Leap Motion の出力を訓練標本として手の動き収集し, 学習を行った. その結果, 精度の高い推定が可能であると確認した.

キーワード: 3次元ユーザインタフェース, 入力デバイス, コンピュータグラフィックス

1. はじめに

人の活動理解やコンピュータアニメーションの生成など, 様々な分野でモーションキャプチャシステムのさらなる発展が望まれている. その中でも手のモーションキャプチャは, 応用用途が広いことから強く期待される分野であるが, 細かな手指の動きは遮蔽が起りやすく光学センサでの計測が難しい. またクレイモデリングや料理などのグローブ型デバイスを利用することができない場面も多く, そのような場面でも利用可能なハンドモーションキャプチャシステムの検討は未だ十分に進んでいない.

本研究では, オクルージョンや自由な手の動きを疎外しない磁気式の無線モーションセンサシステム IM3D[1]を利用して, 手のモーションキャプチャを試みる. IM3D は, LC コイルをワイヤレス・バッテリーレスで識別可能な極小パッシブマーカとし, 同時に最大 15 個までその位置・姿勢 (5 自由度) を 30 Hz 以上の速度で計測できる. このマーカを手手に複数装着すれば, 遮蔽が起りやすく, 手にモノが把持されているような環境でも, それぞれのマーカの情報から手の様子を高精度に推定ことができると考えられるが, 手の代表点全ての位置・姿勢・形状の高精度計測に向けては, 15 点の計測では不十分である. 一般的には, 限られた数の計測点から手の動きを推定するには, 手の各関節の制約条件をふまえて逆運動学を用いる方法が使われる. IM3D でこの方法を用いた場合, モーションセンサの計測値 (32 次元) を計算で位置・姿勢情報 (5 次元) に変換し, その後に手全体の形状・位置・姿勢の推定を行うため, 位置・姿勢計算の段階で利用可能な情報が失

われる可能性がある.

そこで本研究では, より質の高い手のモーションキャプチャに向けて, 従来の逆運動学を用いた計算ではなく, 機械学習したモデルにモーションセンサの計測値を入力し, 手形状・位置・姿勢 (図 1) を推定する方法を提案する. ここで機械学習を積極利用するに至ったのは, IM3D は, マーカからの信号を格子状に配置されたセンサコイルアレイによって計測しており, センサからのデータを画像データと同様の 2次元配列として扱うことができ, 近年画像認識で成功している CNN のモデルによる学習が可能と考えたからである. まずは, 片手に装着した 7 つのマーカに対する IM3D の計測値と光学式のモーションキャプチャの代表例である Leap Motion[4] と OptiTrack[5] の計測値を用いて訓練標本となる入出力のペアを作り, モデルの学習とその評価を行った.

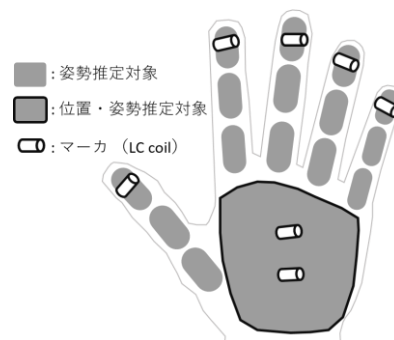


図 1: 推定する手形状

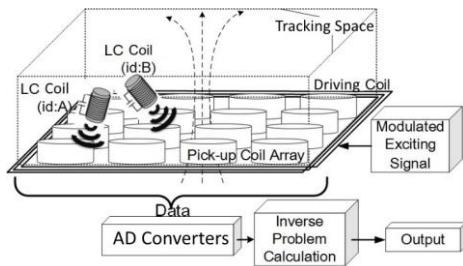


図 2: IM3D のシステム構成と計測原理

2. IM3D: 5DOF 磁気式 3D トラッキングシステム

IM3D は、計測対象である 15 個の LC コイルそれぞれについて位置・姿勢 (5 自由度) の計測が可能である[1]。システムの構成と計測原理を図 2 に示す。詳細な計測方法とハードウェアの構成は、それぞれ参考文献[2, 3]を参照されたい。ここで重要な点は、格子状に並ぶ 32 個のピックアップコイル (各半径 15 mm を 32 個平面状に配置) が得た計測値を位置・姿勢計算に使用することである。計測範囲において、それらピックアップコイルが、LC コイル (各半径約 2 mm, 長さ 15 mm, 重さ約 1 g) が共振して発する磁界を計測して、逆問題を解くことで、それら LC コイルの位置と姿勢を計算する。本研究では、本来 IM3D ではマーカの位置・姿勢の逆問題計算に用いられるピックアップコイルの計測値を機械学習したモデルに入力し、逆問題を解くことなく手形状・位置・姿勢の推定を行う。

3. 手形状・位置・姿勢推定手法

提案手法では、機械学習のモデルにセンサの計測値を入力し、手形状・位置・姿勢を出力する。訓練標本は既存の光学式システムを用いてオクルージョンが発生しない環境で収集するが、学習後は磁気式モーションセンサのみを用いるため、光学的なオクルージョンが発生する環境でも目的の出力が可能になる。この章では、構成したモデルとその学習過程について説明する。なお、学習は 7 つのマーカを用いて片手のみについて行い、モデルのハイパーパラメータは実験的に決定した。

3.1 モデル構成

学習したモデルは、機械学習ライブラリ Keras [6] を TensorFlow [7] をバックエンドとして用い、図 3 のように構成した。IM3D のマーカは、それぞれ共振周波数が異なることで識別可能となっており、使用する 7 個のマーカに対して FFT によって 7 セットのデータが得られる。励起電圧を計測するピックアップコイルは 32 個あるので、モデルへの入力には 7×32 の電圧値となる。前処理では、受けとったピックアップコイルの計測値に対して次の 2 つの処理を行う。

- 1) 2次元配列への変換
- 2) 正規化

32 個のピックアップコイルは、 6×6 の正方格子の 4 隅を欠いた配置をしているため、1) の処理では 4 隅に 0 を追加してデータを正方形の float 型 2次元配列に変換する。

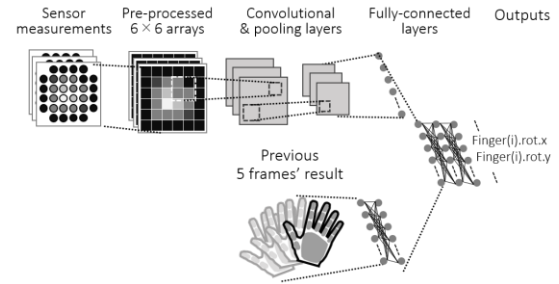


図 3: CNN モデル構成

入力となる電圧値は、マーカが同じ位置・姿勢であっても共振周波数によって電圧の大きさが異なるため、2) では前処理として式(1)を用いて標準化を行う。ここで、 I_i , I'_i はそれぞれ i 番目のピックアップコイルの計測値の標準化前、標準化後のデータを示している。 $E(I_i)$, $\sigma(I_i)$ は、それぞれ順に I_i の平均値、標準偏差を表している。

$$I'_i = \frac{I_i - E(I_i)}{\sigma(I_i)} \quad (1)$$

前処理を経たデータに対し、4 層の畳み込みを行う。層ごとのカーネルの大きさを 2×2 とし、カーネルの数はそれぞれ順に 224, 448, 896, 448 とした。この時、第 1 層のフィルタリングの後、Average Pooling 層を設け、第 2 層の後には Max Pooling 層を設けた。活性化関数には全ての層で *reflected linear unit* (ReLU) を用いた。その後、時間軸方向の情報を加味するため、5 フレーム前までの前処理済みのピックアップコイルの計測値を結合し、3 層の全結合ニューラルネットワークへ入力して、最終的な出力結果を得る。

出力は訓練標本の形式と一致させるため、OptiTrack と Leap Motion によって計測できるデータの形式とする。OptiTrack によって計測した手の代表点の 3 次元位置および姿勢 (四元数) に加え、Leap Motion によって計測した 5 本の指それぞれに姿勢 (オイラー角) を推定する部位は 3 つある (図 1)。指の部位のオイラー角のうち、手の先から二つ目の関節までは自由度が 1 なので、出力の次元数は合計 32 次元となる。事前に全 32 次元を出力するモデルの学習を試みたところ、モデルのハイパーパラメータの値が増大し計算コストが大きくなったため、本研究ではこの 32 次元を手の代表点の位置・姿勢を示す 7 次元 (palm) と各指の部位の姿勢を示す 25 次元 (fingers) に分け、モデル構成は同一に保ったままそれぞれ別に学習した。従って、32 次元の推定のために 2 つの学習済みのモデルを用いる。

学習過程では、学習途中のモデルの出力と訓練標本との差を基に、畳み込み層のフィルタと全結合層の重みを最適化するが、最適化を行う損失関数は Keras で提供される *mean_absolute_error* を用い、最適化アルゴリズムには Adam を用いた。なお、学習率は 0.001 とした。

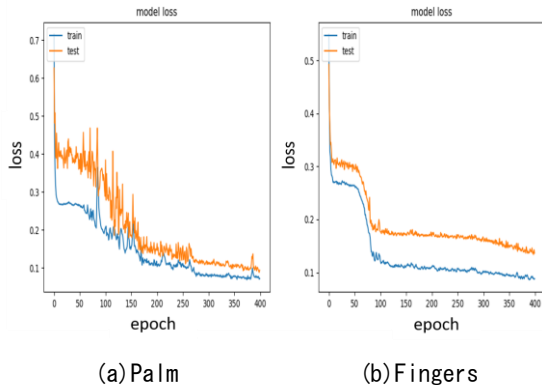


図 4: 学習損失の推移

3.2 学習過程

学習には、ピックアップコイルアレイのある一辺の方向から手を計測範囲に入れ、合計 5120 フレームの訓練標本を収集し、その 10% を学習中にテスト標本として用いた。この学習は、モデルが対象の推定が可能であることを目的としたため、比較的少ない量のデータを用い、計測した手の動作も、手の位置をほぼ動かさず、手を握る・開くという単純な動作のみとした。学習は Nvidia Quadro P2000 でおよそ 15 分かけて行った。訓練標本の収集には、バンド幅の制約から 2 台の Windows PC を用い、1 台を OptiTrack と IM3D の計測値を他方へ送るサーバとし、もう 1 台をそのデータを受け取り Leap Motion の計測値と統合して記録するクライアントとした。

Leap Motion によって得られた指の部位の姿勢を示す訓練標本は、指が動いてオイラー角が減少する際、本来であればが正の値から負の値に転じる時に、 0° から 360° へ値が飛躍する仕様であったため、オイラー角の範囲を -180° から $+180^\circ$ に変換した。また、データの最大・最小値が推定を行う部位によって大きく異なるため、あらかじめ 32 次元の位置・姿勢データの最大・最小値を抽出し、 -1 から 1 の範囲に収まるように式(2)を用いて正規化を行った。ここで O_j 、 O'_j はそれぞれ順に j 次元目の正規化前、正規化後のデータであり、 max_j 、 min_j は j 次元目のデータの最大値と最小値である。

$$O'_j = 2 \frac{O_j - max_j}{max_j - min_j} - 1 \quad (2)$$

4. 学習結果

3 章で示したモデルに対して 400 エポックの学習を行ったところ、損失関数の値は palm および fingers の推定をする 2 つのモデル両方で減少が確認された (図 4)。横軸に示されたエポック数の増加に伴い、縦軸である損失関数の値が訓練標本(train)についてもテスト標本(test)についても減少した。このことから学習が正常に行われたことがわかる。学習を 400 エポック繰り返しても、最後まで微小な減少傾向が見て取れるため、さらに学習が進む可能性がある。

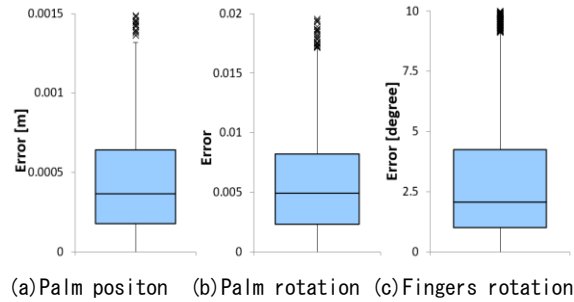


図 5: テストデータ入力時のモデルの出力誤差

モデルの性能を確かめるため、訓練標本を集めるときと同様に評価用のデータを新たに 498 フレーム収集し、学習済みのモデルに入力して、その出力と OptiTrack および Leap Motio の計測値との差を評価した。その際、モデルの出力は式(2)の正規化が行われた後の値の推定値であるため、モデルの出力に逆の操作を施した上で真値と比較したところ、図 5 のようになった (Palm rotation については、真値と推定値のユークリッド距離を Error とし、 -1 から 1 をとる四元数の値で計算したため、単位は存在しない)。図 5 の (a), (b), (c) の中央値はそれぞれ、 0.0004 m, 0.0049 , 2.0654° となり、誤差の分散も小さな範囲に収まった。これにより、モデルが正確な推定が可能であることを示すことができた。誤差の値は、累積のフレーム数が少ないときに大きくなる傾向が見られた。これは、5 フレーム前までの出力結果が得られていないために起きていると推察できる。また、学習時と比較して、損失関数の値は大きかった。これは、訓練標本に含まれていないデータが評価用のデータに含まれていたことが主な原因と考えられる。

5. 議論

5120 フレームのデータによる学習で、単純な手の動きについては良好な結果を示す機械学習のモデルを作ることができた。オクルージョンが発生する状況を作り出して、訓練標本を集めるときに使用した Leap Motion と結果を比較する実験は実施しなかったが、本手法がその状況でも安定した対象の推定が可能なのは原理的に明らかである。

学習およびテストに用いたデータの収集時には、OptiTrack の光学式マーカが手そのものによって遮蔽され、データが収集できないフレームは含まなかった。従って、評価結果には表れていないが、実際には推定できない姿勢が存在する。しかし、これは OptiTrack のカメラの配置を変えながら標本を増やすことで対応可能である。

IM3D の現在の同時利用可能なマーカの数は 15 個までなので、両手について本手法を適用することも可能である。しかし、学習・テストデータの収集は片手のみについて行い、ピックアップコイルアレイのある一辺の方向から手を計測範囲に入れて行ったため、もう片方の手と指にマーカを付けた場合や、ピックアップコイルの別の辺など他の方向から手を計測範囲に入れた場合には、適切な推定結果が得られない可能性が高い。両手にマーカを付けた場合や、

様々な方向から手を入れて訓練標本を収集すれば解決できるが、標本収集のための時間的コストが増大することになる。この問題については、IM3Dのピックアップコイルアレイの対称性および手の対称性を用いてコストの削減をすることで対応できる。実際には、データを取得するデバイスの座標系の原点をピックアップコイルのある平面の中心とし、手を入れたピックアップコイルアレイの辺の法線方向を軸として、すでに収集済みの訓練標本のデータを反転すれば、もう一方の手の訓練標本として利用できる。また、ピックアップコイルの中心を軸に 90° ずつデータを水平方向に回転させるように計算すれば、様々な辺から手を計測範囲に入れた時と同等の訓練標本が生成できる。

6. 結論

本稿では、オクルージョンが発生し、かつグローブ型デバイスが使用できない環境でも手形状・位置・姿勢の推定を可能にすべく、オクルージョンを問題としないIM3Dの磁気式モーションセンサの計測値を入力とし、対象の推定を行うCNNのモデルを提案した。OptiTrackとLeap Motionで集めた5120フレームの訓練標本を使用してモデルを学習し、その推定誤差は、手の代表点の位置・姿勢および指の部位の姿勢に対して中央値がそれぞれ順に0.0004 m, 0.0049, 2.0654° となり、高い性能を示した。これにより、訓練標本を増やすことで、より複雑な動きでも形状・位置・姿勢の推定ができる可能性を示すことができた。

謝辞 本研究の一部は JSPS 科研費 (18H04103) による。

参考文献

- [1] Huang, J., Takashima, K., Hashi, S., and Kitamura, Y.: IM3D: Magnetic motion tracking system for dexterous 3D interactions. *SIGGRAPH Emerging Technologies*, 12:1, 2014.
- [2] Hashi, S., Toyoda, M., Yabukami, S., Ishiyama, K., Okazaki, Y., Arai, K., and Kanetaka, H.: Wireless magnetic motion capture system using multiple LC resonant magnetic markers with high accuracy. *Sensors and Actuators A: Physical* 142, 2, 520–527, 2008.
- [3] 藪上信, 栢修一郎, 徳永裕樹, 河野丈志, 荒井賢一, 岡崎靖雄: バッテリおよび給電線を持たない LC 共振型磁気マーカの位置検出システム. *Journal of the Magnetics Society of Japan* 28, 877–885, 2004.
- [4] OptiTrack. <http://optitrack.com/>
- [5] Leap Motion. <https://www.leapmotion.com/>
- [6] Keras. <https://keras.io/>
- [7] TensorFlow. <https://www.tensorflow.org/>