

関節回転制約を考慮したメッシュ回帰による手指姿勢推定の検討 Hand Reconstruction using Mesh Regression with Joint Rotation Constraints

村上 智哉[†]

Tomoya Murakami

箱崎 浩平[†]

Kouhei Hakozaki

高橋 正樹[†]

Masaki Takahashi

金子 浩之[†]

Hiroyuki Kaneko

1. はじめに

手話の実写映像から人物の動きを手話モーションとして効率的に収集する技術の研究を進めている。深層学習を用いた手指姿勢推定モデルを活用することで、カメラ 1 台の簡易な環境から 3 次元の動作情報を取得し、効率的な手話モーションの収録が期待できる。

手話では手指の位置や指先の接触など細かな動きが意味を表す重要な要素になる。深層学習を用いた手指姿勢推定の既存手法では、手指同士の遮蔽により推定される形状が破綻し、そこから得られる手指の姿勢も不自然となる課題があった。本稿では、手指形状の破綻を軽減させるために、メッシュ回帰モデルに関節回転制約を導入する手法を提案する。実験により、関節回転制約導入前と比較し推定精度が向上するとともに、遮蔽の伴う入力画像に対して手指形状の破綻が軽減されることを確認したので報告する。

2. 関連研究

実写映像に基づく手指の姿勢解析は、VR や AR など多くの 3D アプリケーションで活用されている。大規模なハンドデータセット InterHand2.6m[1]が公開されて以来、単眼の RGB 画像から両手指の姿勢を推定する手法がいくつか提案された。これらの手法は大きく分けて、パラメトリック[2]とノンパラメトリック[3,4]の 2つのアプローチに分類される。

パラメトリックなアプローチは、MANO[5]に代表されるパラメータ制御可能なハンドモデルを駆動することを目的としている。画像から、ハンドモデルの指の曲がり具合を制御するポーズパラメータと、手の厚み・長さを制御する形状パラメータを回帰し、その情報から事前に定義されたパラメータ空間でハンドモデルを制御する。これにより、もっともらしい手指の姿勢を得ることができる利点がある。しかしながら、パラメータ空間外の姿勢は表現できず、入力画像の手指の姿勢を完全に再現することはできない[6]。

ノンパラメトリックなアプローチは、画像から手指の皮膚表面に定義されたメッシュの 3 次元頂点座標を回帰し、手指形状をモデルから直接得ることができる。このアプローチは、パラメトリックな手法と比べ、精度良く手指の姿勢を推定できることが報告されている[3]。代表的な手法である IntagHand[4]は、メッシュトポロジーを明示的に符号化するためのグラフ畳み込みネットワークと、メッシュ頂点の関係性をモデル化するための多層トランスフォーマーエンコーダーを組み合わせ、これまでの手指姿勢推定手法を大きく上回る性能を達成した。一方で、ノンパラメトリックなアプローチは、手指の重なりで過度な遮蔽が生じた際、遮蔽された指が不自然な方向へ曲がるなど、人間の手

[†] NHK 放送技術研究所 NHK Science & Technology Research Laboratories

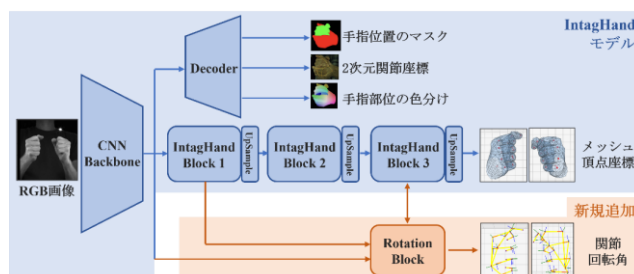


図 1 提案するモデル構造

指として破綻した形状を推定する場合がある。

本稿ではノンパラメトリックなアプローチの課題解決の為、回転方向や可動域など手指関節構造に基づいた制約を、深層学習モデルに学習させることを提案する。これにより、ノンパラメトリックなアプローチの長所である高い精度を活かしつつ、手指形状の破綻を軽減した推定が可能となる。

3. 提案手法

3.1 モデル構造

図 1 に提案するモデル構造を示す。関節回転角を推定するため、IntagHand[4]で提案された IntagHand Block をベースに、座標-回転情報の Cross-Attention 機構を加えた Rotation Block を追加した。これにより、入力された RGB 画像からメッシュ 3 次元頂点座標 $V \in \mathbb{R}^{N_v \times 3}$ と関節回転角 $R \in \mathbb{R}^{N_r \times 6}$ が片手毎に出力される。ここで N_v はメッシュ頂点数、 N_r は回転可能な手指関節数を表す。関節回転角は Zhou らの報告[7]を参考に 6 次元回転表現を用いている。回転角は取りうる値の範囲が可動範囲としておおよそ決まっているため、これを学習データに含むことで、遮蔽のある画像でも可動域外の角度を推定しないようにモデルが学習することを期待する。

今回、既存手法との比較の為、 $N_v = 778$ からなる MANO のメッシュトポロジーを採用し、回転可能な手指関節数は、各指の第 1 ~ 3 関節に手首のグローバル回転を加えた $N_r = 16$ とした。

3.2 損失関数

回転角を推定するネットワークを追加したことに伴い、[4]で使われている損失関数に 3つの損失関数を追加した。

Rotation Loss

回転角の推定値と正解データを近づけるため、L1 損失を使用した。

$$L_{\text{rotate}} = \sum_{i=1}^{N_r} \|R_i - R_i^{GT}\|_1$$

R_i は i 番目の関節回転角の推定値、 R_i^{GT} はその正解データをそれぞれ示す。

Non-rotatable Euler Loss

手指関節は構造上、回転可能な方向として DoF (Degree of Freedom) が各関節に存在する[8]。本来回転しない方向に指が曲がることは見かけ上の違和感となる。モデルがこのような関節回転制約を効率的に学習できるようにするため、指が曲がらない回転方向をオイラー角として抽出し、回転方向に制約をかけるための関数を L2 損失で実装した。

$$L_{unable} = \sum_{i=1}^{N_r} \sum_{E_{i,j} \in E_{unable}} \|E_{i,j}\|_2^2$$

ここで、 E_{unable} は指が曲がらない方向のオイラー角群、 $E_{i,j}$ は i 番目の回転可能な手指関節における j 成分 ($j \in x, y, z$) のオイラー角を示す。ただし $E_{i,j}$ は関節回転角の推定値 R_i から算出した。

Position-Rotation Consistency Loss

IntagHand Block と Rotation Block の出力から得られる手指の姿勢は互いに一致していることが望ましい。そこで、それぞれの出力である座標と回転角から、手指を構成する骨の向きを算出し、整合をとるため関数を L2 損失で実装した。なお、ここでの骨の向きとはグローバル座標系における、各手指関節から隣接する指先方向の関節 (子関節) へ向けた 3 次元単位ベクトルを指す。

$$L_{consist} = \sum_{i=1}^{N_b} \|\mathcal{D}_{pos}(J_i, J_{c,i}) - \mathcal{D}_{rot}(R_i)\|_2^2$$

ここで、 N_b は骨の数、 J_i は i 番目の関節の 3 次元座標、 $J_{c,i}$ はその子関節の 3 次元座標、 \mathcal{D}_{pos} と \mathcal{D}_{rot} はそれぞれ関節座標と回転角から骨の向きを求める関数である。ただし 3 次元関節座標 J は、あらかじめ定義された回帰行列を用いて、3 次元メッシュ頂点座標 V から算出した。

4. 実験

RGB 画像と両手メッシュのアノテーションを持つ InterHand2.6M[1]に、関節回転角を追加アノテーションした上で学習データセットとして利用した。入力画像に対し IntagHand[4]と同様の前処理を行ったうえで、366K の学習用データと 261K のテスト用データに分割した。[4]をベースラインとし、バッチサイズ、学習率など同条件で学習を行い、最も精度が高かったモデルを評価に用いた。

表 1 に提案手法と [4] のメッシュ頂点と関節の座標誤差を比較した結果を示す。評価手法には、それぞれ MPVPE (Mean Per Vertex Position Error) と MPJPE (Mean Per Joint Position Error) を用いた。図 2 ではテストセットにおけるメッシュ頂点座標の正解データとの誤差をもとに、0 ~ 50 の閾値での正解率をグラフにした 3D PCK (Percentage of Correct Keypoints) 曲線と AUC (Area Under the Curve) を示す。表 1、図 2 より、提案手法は全ての評価方法でベースラインを上回っており、関節回転角の回帰部をモデルに組み込むことで、推定されるメッシュ頂点座標の精度が向上することを確認した。

図 3 に遮蔽を伴う画像を入力とした手指形状の推定結果を示す。[4]では正面・俯瞰ともに遮蔽された右手の形状がつぶれているが、提案手法では破綻なく手指形状の推定ができています。これにより、関節の回転制約が、遮蔽の伴う入力画像に対して手指形状の破綻を軽減した推定を行うために有効であることが示唆された。

表 2 関節・メッシュ頂点誤差の比較

手法	回帰対象	MPVPE ↓	MPJPE ↓
ベースライン	メッシュ	9.73	9.52
IntagHand[4]			
提案手法	メッシュ+回転角	9.33	9.12

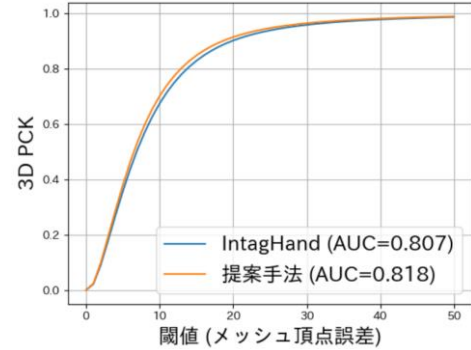


図 2 3D PCK および AUC の比較

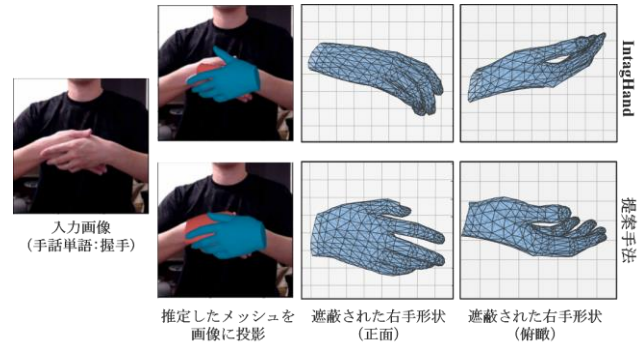


図 3 遮蔽された手指形状の定性的な比較

5. おわりに

破綻の無い手指形状を推定することを目的とし、関節回転制約を考慮したメッシュ回帰モデルを提案した。実験によりベースラインと比較し精度よく手指姿勢を推定できるのを確認したとともに、遮蔽の伴う入力画像に対して手指形状の破綻を軽減した推定ができることを示した。

参考文献

- [1] G. Moon, S. Yu, H. Wen, T. Shiratori, K. Lee, "Interhand2.6m: A dataset and baseline for 3d inter-acting hand pose estimation from a single rgb image", Proc. ECCV, pp.548-564 (2020)
- [2] B. Zhang, Y. Wang, X. Deng, Y. Zhang, P. Tan, C. Ma, H. Wang, "Interacting two-hand 3d pose and shape reconstruction from single color image", Proc. ICCV, pp.11354-11363 (2021)
- [3] M. Li, L. An, H. Zhang, L. Wu, F. Chen, T. Yu, Y. Liu, "Interacting Attention Graph for Single Image Two-Hand Reconstruction", Proc. CVPR, pp.2761-2770 (2022)
- [4] K. Lin, L. Wang, Z. Liu, "Mesh Graphormer", Proc. ICCV, pp.12939-12948 (2021)
- [5] J. Romero, D. Tzionaz, M. Black, "Embodied hands: Modeling and capturing hands and bodies together", ACM Trans. Graph, 36, 6 (2017)
- [6] C. Wang, F. Zhu, S. Went, "McMaHand: Exploiting Mesh-Mano Interaction for Single Image Two-Hand Reconstruction", Proc. CVPR, pp.564-573 (2023)
- [7] Y. Zhou, C. Barnes, J. lu, J. Yang, H. Li, "On the Continuity of Rotation Representations in Neural Networks", Proc. CVPR, pp.5738-5746 (2019)
- [8] 井樋栄二, 吉川秀樹, 津村弘, 田中栄, 高木理彰, "標準整形外科学 第 14 版", 医学書院 (2020)