

MR 環境下における身体背面部への タップ入力を用いた入力領域拡張

國枝 直希^{1,a)} 真鍋 宏幸^{1,b)}

概要：ヘッドセットや AR グラスなどの普及により，ユーザの手を用いて自然なインタラクションを可能にするハンドジェスチャが注目されている．ハンドジェスチャを用いれば，手でオブジェクトや UI を操作することができるが，HMD の内蔵カメラに手が映る範囲でしか利用できない問題や入力語彙が少ない問題がある．それらの課題に対し本稿では，内蔵カメラと手首に装着した IMU を用いた，身体背面部を含む様々な部位へのタップ入力手法を提案する．提案手法の有用性を確認するため，立位，座位の 2 種類の姿勢に対して腰や後頭部などの身体背面部へのタップを検出する精度を評価する実験を実施した．その結果，左右各 4 か所ずつに対して立位では平均 76%，座位では平均 78%の精度が得られた．

1. はじめに

ヘッドマウントディスプレイ (HMD) の技術進化に伴い，複合現実 (MR) などの分野が急速に発展し，多岐にわたる応用が進んでいる．エンタテインメント，教育，医療，訓練シミュレーションなど様々な分野でその可能性が探求されており，特に，没入感や臨場感の向上がユーザ体験の改善につながることから，それらを高めるようなさまざまなインタラクション手法が研究されている．

現在の VR/MR アプリケーションで用いられているインタラクション手法の一つとして，ハンドジェスチャがある．この手法では，ユーザ自身の手の動きを利用して仮想環境とのインタラクションを行っており，没入感を損なうことなく自然な操作性を実現している．しかし，ハンドジェスチャには制約もある．ハンドジェスチャによる入力は通常，HMD に搭載されたカメラを利用してユーザの手の動きを認識することで実現しているが，VR/MR 向け HMD の内蔵カメラの視野には制限があるため，操作に利用できる領域が HMD の前方側に限られてしまう問題がある．また，ジェスチャによる入力語彙を拡張するために，認識するジェスチャの種類を増やす方法がある．ジェスチャの種類を増やす場合，識別するジェスチャの種類が増えるほど，ユーザの意図しない動作に対して操作が行われてしまう問題 (ミダスタッチ問題) や，ジェスチャの複雑化によるユーザ負担が増大する問題が発生するため，現実的な入力語彙

の上限に制約がある．別の拡張方法として，ジェスチャを行った場所や手の向いている方向などをもとに入力の意味を変える方法がある．例えば，仮想キーボードは，選択された場所によってキーが異なるため，多くのパターンを入力できる．しかし，仮想キーボードを利用した場合，視野の一部がキーボードの UI で覆われる点や，見ないで入力することができない点に課題がある．

このような制約を緩和する手法として，ユーザの身体表面を入力領域として用いる On-Body インタラクションに注目が集まっている．On-Body インタラクションでは，身体上の特定の部位をタップするなど，ユーザの身体を入力インタフェースとして利用している．これにより，ユーザは自分の身体という通常のデバイスよりはるかに大きな領域をウェアラブルデバイスとして利用することができ，自然でより自由な入力を行うことができる [1]．また，固有感覚を活用することで，ユーザは目を閉じていても自分の身体の位置や動きを感じ取ることができるため，操作領域を見ることなく利用することができる．このような利点から，身体表面を入力領域として利用する研究が数多く行われているが，これらの研究であっても，HMD の内蔵カメラの視野範囲内や，センサを装着した身体部位，およびその延長線上でのインタラクションにとどまっており，入力範囲に制約がある．

そこで，本研究では，VR/MR 環境での入力語彙を拡張することを目的に，身体の後方や背面部など，これまで利用されてこなかった領域にまでタップ入力領域を拡張する手法を提案する．

¹ 芝浦工業大学

^{a)} al20082@shibaura-it.ac.jp

^{b)} manabehiroyuki@acm.org

2. 関連研究

本研究は、ジェスチャとそれを実現するためのトラッキング技術、身体を入力デバイスとして用いることで入力語彙を拡張する研究などに関連する。

2.1 トラッキングとジェスチャ

2.1.1 ウェアラブルなボディトラッキングに関する研究

ボディトラッキングは、ユーザの身体の動きをリアルタイムに反映することで、VR/MR 空間上でユーザと自然に対話するための手法として重要な役割を果たしている。

これまで、カメラベース、逆運動学ベース、機械学習ベースなど、様々なトラッキング手法が研究されている。例えば、Cao らは、カメラで撮影した画像、動画内の人物に対してリアルタイムに姿勢推定を行うシステムを提案している [2]。このシステムを利用することで、カメラの情報のみでボディトラッキングを行うことができるが、データセット内に無い姿勢での推論失敗や部位の欠落、写り込みによる偽陽性などの課題がある。また、ユーザの全身の動きが隠れることなくトラッキングできるようにするには、カメラの死角がないように複数のカメラを設置する必要がある。

Caserman らは、HTC Vive と Vive Tracker を用いることで低遅延、高精度のボディトラッキングを行っている [3]。この方法では両手首と両足首の計 4 か所に Vive Tracker を装着し、逆運動学を少ない反復回数と計算時間で効率的に解くことで、高い精度と低遅延を実現した。しかし、この方法では両手首と両足首の計 4 か所に Vive Tracker を装着し、Vive Tracker を認識するためのベースステーションを設置する必要があるため、屋外などベースステーションの設置が難しい環境では利用できない課題が残る。

カメラやトラッカーなどの外部機器を用いずに全身トラッキングを行う研究として、HMD と左右の手に持ったコントローラの 3 点の情報から強化学習によってユーザのアバタ全身の姿勢を推定する研究が行われている [4][5]。これらの研究では、頭部と両手の 3 点の座標や姿勢情報のみで全身の動きを再現することができるものの、上半身と下半身の動きの相関が弱い場合に、ユーザが同じような上半身の姿勢をしても違う下半身の姿勢情報の出力が得られるなど、同じ動作を行ってもユーザの意図しない姿勢となる問題や、学習外の動きをした際に大きく姿勢が崩れる可能性などの課題がある。

2.1.2 ハンドトラッキングとジェスチャに関する研究

Chiu らは、Leap Motion とアクションカメラ用のジンバルを組み合わせることで、ハンドトラッキングを行える空間を拡張する方法を提案した [6]。この研究では、身体に装着した自己駆動式のジンバルに Leap Motion などのカメラセンサを取り付け、ユーザの手の位置に合わせて動的に

方向を変えることでハンドトラッキングを行える空間を拡張し、より広い範囲でジェスチャを行っている。しかし、身体背面部などの領域でのハンドトラッキングは試みられていない。

Shen らは、スマートウォッチに搭載されたモーションセンサと地磁気センサ (9 軸 IMU) を用いて、腕全体 (肘、手首) の動きを推定する手法を提案した [7]。この研究では、腕の解剖学的モデルによる制約と、腕全体の姿勢が固定された肩の位置を基準とした肘の位置、手首の位置、手首の回転の 3 つのパラメータで一意に定まることを利用して、9 軸 IMU の値をもとに肩から見た相対的な手首の位置を推定している。この手法は、機械学習に依存しない手法としては高精度のトラッキングを行うことができたが、ユーザの移動や方向転換を考慮していないことや、計算速度などに課題がある。

Wei らは、スマートウォッチに搭載された 6 軸 IMU を利用して肘、手首の位置を RNN によって推論する手法を提案している [8]。この手法では、肩を原点とした座標系に対し、RNN を用いて手首と肘の回転を推論し、その値と前腕、上腕の長さをもとに腕の可動域の制約に従って肩との相対位置として肘の位置を、肘との相対位置として手首の位置を計算している。この研究では 9 軸 IMU (加速度、ジャイロ、地磁気) や、外部デバイスを利用しなくともトラッキングが行えることを示したが、肩や胴体が静止している仮定の下で、肘との相対位置として手首の位置を推論しているため、肘の誤差の影響により手首の絶対位置の誤差が大きくなる問題や、肩が静止した状況下でしか利用できない制約がある。

このように、カメラや IMU を用いてハンドトラッキングを行う手法が盛んに研究されているが、カメラを用いた手法は精度が高いものの遮蔽や死角に弱く、IMU を用いた手法は遮蔽に強くノイズによる誤差の累積に弱い点がある。このように、カメラベースの手法と IMU ベースの手法にはそれぞれ利点と欠点があるが、それらを組み合わせることで互いを補完しあう手法も研究されている。Streli らは、VR 空間上で HMD のハンドトラッキング範囲外のオブジェクトに対して手による操作 (つかむ、引っ張る、移動させるなど) を行う HOOV を提案した [9]。この研究では、HMD に搭載されたカメラでのトラッキング範囲内ではカメラによって、範囲外でのトラッキングは過去のトラッキング情報や手首に装着した IMU のセンサ情報などから機械学習モデルによって手首の位置や姿勢を推論することで、6 軸 IMU でも高い精度で身体後方でのハンドトラッキングを行っている。ハンドトラッキングと同時に IMU の値からピンチジェスチャを検出することで、内蔵カメラの視野外の領域もピンチジェスチャが可能な領域へと拡張した。本研究では、同様のハンドトラッキング手法を用いるが、HOOV のような身体後方へのピンチジェスチャ

の範囲の拡張ではなく、タップを検出できる領域の拡張を行い、全身の表面へ適用する。

2.2 身体を入力デバイスとして用いる研究

身体表面を入力デバイスとして用いる手法は、身体が常に利用可能なウェアラブルデバイスであること [10]、視覚フィードバックに依存しない多様な入力を行うことができるアイズフリーな入力手法であること [11]、ユーザの身体を用いることにより自然に操作できることなどの利点から、数多くの研究が行われている。例えば、Harrison らは、腕をタップした際の振動を、センサを搭載したアームバンドで検知することで、腕を入力デバイスとして用いる Skinput を提案している [12]。Skinput では、体内に伝わる振動が周波数帯域によって異なる伝搬特性を持つことを利用し、指で手や腕をタップした際の振動の振幅と周波数の分布を特徴量として抽出するとともに、どの位置をタップしたのかを機械学習の手法で分類することで、肌の表面を入力領域として利用している。しかし、この手法では、アームバンドを装着した側の腕や手しか入力領域に用いることができない。

また、Mi らは、座位での HMD 利用時に膝をタッチインターフェースとして使用する LapTouch を提案した [13]。LapTouch では、膝と手の位置を OptiTrack でトラッキングし、膝上の領域をタッチ可能な格子状の領域に分割することで、座った状態での VR 環境下の入力を拡張している。それに対し、本研究では、HMD と入手の容易な 6 軸 IMU のみを用いて立った状態、座った状態のどちらでも利用できる手法を実現する。

全身を入力領域とした際にどの程度のパターン数に分割するのが実用的なのかを調査した研究も行われている。Cheng らは、VR 環境下において、トラッキングスーツを用いて全身を入力領域とした際に、12, 20, 28, 36 個の異なる領域に分割した際のタッチ精度に関して調査した [14]。その結果、12, 20 個のボタン領域に分割した際は 95% の精度でタッチを行うことができたが、それ以上の数に分割した際には一般化されたタッチパターンを利用しても精度が上がらず、個人のタッチ習慣に大きな違いがあることが示唆された。このように、領域を細かく分けたとしてもユーザ側のタッチ習慣や、ユーザが見なくても認識できる細かさによる限界があることから、本研究では、HMD と左右の手首に装着する IMU といった最小限のデバイスで全身を入力領域にし、大まかな部位に分割することで実用上問題のないパターン数での実現を目指す。

3. 提案手法

本研究では、ボディトラッキング、内蔵カメラを利用した手首のトラッキング、および IMU によって、タップされた身体の部位を検出し、部位と対応する操作（カーソル

移動、決定など）を行うことで、手でタップ可能な身体の部位を入力領域として利用するシステムを提案する。

本手法では、HOOV[9] と同様に、左右の手首に IMU を装着し、HMD によるトラッキング範囲内では HMD、範囲外では過去のトラッキング情報や IMU によって手首のトラッキングを行う。推論された手首の位置と姿勢情報から、手のひらの位置を計算する。トラッキングと平行してタップの検出やタップした部位の推定を行い、部位に対応した操作を事前に割り当てることでタップジェスチャによる操作を可能にする。

提案手法を用いて腰の右側をタップしている様子を図 1 に示す。また、図 2 に今回検出対象とする部位（青丸）と、提案手法でカバー可能な他の部位の例（黒丸）を示す。図に示すように、本研究では身体背面部（後頭部、肩、背中、腰）へのタップ操作を対象とするものの、提案手法自体は、手首のトラッキングによりタップ操作が可能なすべての部位に適用可能である。したがって、提案手法は、最終的に全身を入力デバイスにすることができる手法となることが期待される。

4. 実装

4.1 システム構成

ユーザの左右の手首に装着する IMU センサにはリストバンドに装着した Joy-Con を、HMD には Meta Quest 2 を用いて実装を行った。

Quest 2 向けの VR アプリケーションの実装は Unity 2021 で行い、ソケット通信を介して Python で作成したアプリケーションと VR アプリケーション間で情報を送受信している。Python アプリケーションは、Core i9-9900K, GTX 960, 32GB RAM を搭載した PC 上で動作している。

4.2 機械学習モデル

手首の位置と姿勢を推定する機械学習モデルについては関連研究で示した HOOV[9] のモデルと同様のものを PyTorch で実装する。パラメータの初期化については He の手法 [15] を用い、最適化関数は Adam[16]、損失関数については、位置と姿勢それぞれの L1Loss の和に設定する。カルマンフィルタやその非線形拡張についてまとめられた研究 [17] を参考に、手首の姿勢を推定する拡張カルマンフィルタを実装した。また、入力に用いる姿勢情報のパラメータについては、HOOV と同様に、Meta Quest 2 から得られた過去 5 回分の頭、手の座標/姿勢情報、手が内蔵カメラの視野から外れた後の頭の座標/姿勢情報、加速度/ジャイロセンサの値、拡張カルマンフィルタの出力した手首の姿勢情報を用いた。姿勢表現を回転行列の第 1, 2 行を連結した 6 次元表現に変換を行うことで、回転表現の連続性が保証され、ニューラルネットの頑健性が向上することが知られている [18] ため、先行研究 [4][5][9] と同様に、入力パラ



図 1 提案手法. HMD を装着したユーザが腰をタップする様子.

メータ中のすべての姿勢情報を 6 次元表現に変換を行う. 学習に用いるデータセットについては, Vive Tracker 3.0 を頭部及び左右の手首に装着し, 頭部と左右の手首の相対的な 3 次元座標と姿勢情報を取得することで作成した. 手が内蔵カメラの視野内であっても Joy-Con を装着したリストバンドやその他の原因による遮蔽が起きた際に, その近傍のデータが学習されていないと機械学習モデルの出力が不安定になる可能性がある. そのため, 遮蔽が起きた際でも安定的に推論できるように, データセット作成の際には内蔵カメラのトラッキング範囲内でのデータも含めた. また, 今回の提案手法に合わせたデータセットを作成するため, 身体の各部位をタップしながらデータセットを作成した. 最終的に, データセットのデータ数は, 片手あたり約 18,000 件 (10 分～15 分に相当する件数) となった. また, 出力に関して, 位置については学習データ取得時の正解位置と, 実際に利用する際の位置がユーザの移動などによって定数分ずれている可能性を考慮し, HMD の内蔵カメラの視界から手が離れる直前の座標に予測値の変化量を加える形で推論値を決定する. すなわち, カメラの視界から手が離れる前の最後の手首の座標を p_{last} , 手が視界を離れて最初の予測値を \hat{p}_{start} , それ以降の予測値を \hat{p} とおくと, 補正された予測値 $\hat{p}_{corrected}$ は,

$$\hat{p}_{corrected} = p_{last} + (\hat{p} - \hat{p}_{start})$$

のように計算する. また, すべての入力, 出力の座標系を Unity で用いられている左手座標系に変換した.

4.3 各部位の位置の推定

身体の各部位の座標については, 立位状態での HMD の Unity 上の座標をもとにユーザの身長を算出し, 身長をもとに統計的な数値からユーザの体格と各部位の相対的な位置を設定した. 統計データに関しては, 身長は e-Stat の 2019 年の 20 歳～29 歳のデータ^{*1}を, その他の部位に関し

^{*1} <https://www.e-stat.go.jp/dbview?sid=0003224177>

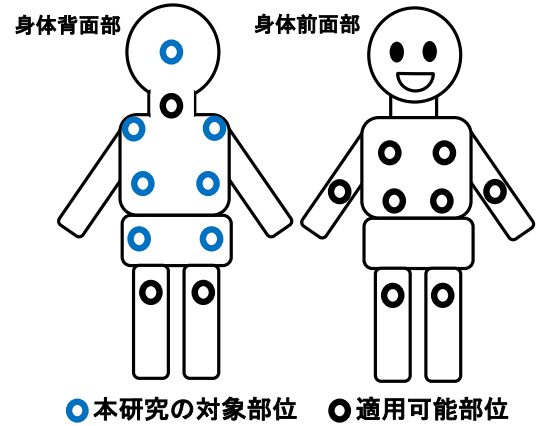


図 2 検出を行う部位

ては, 2014～2016 年に計測された日本人男性のデータのうち, 25 歳以下に關してまとめられたデータ [19] を参考にしている. この手法では, HMD から得られる頭部の情報のみで身体部位の位置推定を平易な計算で行うため, 精度が低下する可能性はあるが, 追加の機器を装着することなく, 必要な計算リソースも少なくできる上, 座るなどの動作によって姿勢が大きく崩れる可能性を考える必要がない.

次に, 各部位の相対的な位置の計算方法について述べる. 身長の高さ H の分布と身体を構成する各部位の大きさの分布がどちらも正規分布に近似できると仮定する. この時, 頭部の高さを y_{hmd} , 計算された身長を $\hat{H} = y_{hmd}/0.95$, 身長の平均値を μ_H , 標準偏差を σ_H , 身体の各部位の高さ, 長さなどの平均値を μ_{body} , 標準偏差を σ_{body} とするとき, 身長 \hat{H} に対する各部位の高さ, 長さの値を \hat{x}_{body} とおくと,

$$\hat{x}_{body} = \mu_{body} + \frac{\hat{H} - \mu_H}{\sigma_H} \times \sigma_{body}$$

のように計算できる. この計算式及び, 各部位の高さ, 横幅などの各種統計値 (直接の値が存在しない場合は付近の部位の値から計算することで近似した値) から, ユーザの各部位の位置の計算を行った. また, 座位姿勢においても位置が計算できるように, 立位状態での身長を計算値を保存し, 立位状態での位置から身長を計算値の変化分 $\Delta\hat{H}$ を各部位の y 座標から減算した.

4.4 タップ検出および部位推定について

タップ検出は, Joy-Con に搭載された IMU の 3 軸加速度センサから得られた 3 次元ベクトル $\mathbf{a} = (a_x, a_y, a_z)$ のノルム $\|\mathbf{a}\|$ に対して閾値を設定して検出する. 今回の実装では, 閾値を $\|\mathbf{a}\| = 30.0 \text{ m/s}^2$ に設定した. タップした部位の推定は, タップ検出を行った際に計算された手のひらの位置と最も近くにある部位を推定することで行う. 手のひらの位置 $\hat{p}_{palm} \in \mathbb{R}^3$ の計算については, Quest2, または機械学習モデルから出力された最新の手首の位置 $\hat{p}_{wrist} \in \mathbb{R}^3$,



図 3 練習用アプリケーションの画面（図中央：練習する部位，図下：検出された部位）

手首の姿勢に対応する回転行列 $\hat{R}_{wrist} \in \mathbb{R}^{3 \times 3}$ を用いて，

$$\hat{p}_{palm} = \hat{R}_{wrist} (\pm 0.03, 0, 0)^T + \hat{p}_{wrist}$$

のように， x 軸方向に 0.03 m のオフセットを適用することで計算する．なお，0.03 の符号については，左手の時に正，右手の時に負となる．

5. 評価実験

提案手法の有用性を評価するため，2 種類の姿勢でタップ検出精度とタップした部位の推定精度を測定する実験を実施した．実験はアンケートも合わせて約 50～60 分間であり，実験参加者は体格の異なる 20 代男性 5 名（A～E）である．また，参加者には実験に協力した報酬として Amazon ギフト券 1,000 円分を渡した．

5.1 実験手順

5.1.1 実験環境

実験では，実装に用いたものと同じデバイスを用いた．また，Meta Quest 2 を USB Type-C ケーブルを用いて Quest Link によって 4 節で用いた PC と接続し，実験アプリケーションを実行した．

5.1.2 実験で計測する対象部位

2 つの実験において，実験参加者は，左手で後頭部，右肩，背中の左側，左腰を，右手で後頭部，左肩，背中の右側，右腰をタップする．

5.1.3 実験前練習

実験参加者はそれぞれの実験を行う前に，指定された部位をタップする練習を行う．右手で後頭部のタップ練習を行った際の練習用アプリケーションのディスプレイ表示を図 3 に示す．このアプリケーションでは，ユーザがタップ練習を行う部位が画面上に表示され，実際にタップを行うと，フラッシュエフェクトの後にタップに用いた手と検出された部位が表示される．参加者はこの表示を元に練習を行い，十分に練習を行ったと感じた時点で次の部位の練習へ移行し，すべての部位の練習を完了した後に実験を開始する．

5.1.4 実験 1: 立位状態

実験 1 では，参加者は提案手法を用いて立位状態で，事

前に指定した方の手で 5.1.2 節で述べた手と対応する 4 か所の指定部位（後頭部，肩，背中，腰）をランダムに各 10 回ずつタップするタスクを行い，その後もう一方の手で同様のタスクを行う．実験では，参加者は 1 回ごとにタップを行う手が視界に入った状態（HMD の内蔵カメラによるトラッキングが機能する状態）から動作を開始する．実験の間，参加者は立位姿勢を維持する．

5.1.5 実験 2: 座位状態

座った状態などの理想的でない姿勢においてもタップ検出が行えるかを確認するため，座位姿勢で同じタスクを行う．参加者の身長と体格を推定するため，タスク開始前に参加者は立位姿勢を取り，その後座位へと移行する．なお，座った際に背後に手を移動させる空間が空くようにするため，参加者には椅子に浅く座るように指示を行う．

5.1.6 アンケート

実験終了後，実験参加者にいくつかの質問を行う．質問の内容を以下に示す．

- (1) 手首にデバイスを装着することによる手の動かしにくさが生じたか
- (2) 実験を通してタップを続けた結果，痛くなった場所はないか
- (3) 実験 1, 2 の中で，タップがしにくい部位はあったか
- (4) タップした部位に割り当てたい操作や利用したいアプリケーションの例
- (5) 実験全体を通じて気になったことはあるか

なお，これらの質問はすべて口頭で行い，自由に回答を行っていいものとした．

6. 結果

6.1 実験 1 について

実験 1 での実験参加者全員（A～E）の部位ごとの検出精度およびタップ検出精度を表 1, 2 に示す．表に示した結果では，タップ検出そのものに失敗した場合は推定を失敗したものとして計算している．実験 1 全体を通してのタップ検出精度は平均 99.5%，部位推定精度は平均 76.25% であった．今回の実験では，タップ検出に関してはほぼすべてのタップが検出された．部位推定精度については，後頭部と肩については個人差も少なく，高い推定精度が得られた．しかし，背中と腰については，推定精度の高い実験参加者もいれば，ほとんど推定に失敗する実験参加者もいるなど，個人差の大きい結果となった．

また，誤検出された部位を整理するため，混同行列を表 3, 4 に示す．混同行列を見ると，肩に関しては概ね後頭部との誤検出が多く，背中と腰については互いに誤認識しやすい傾向にあることがわかる．しかし，特に右手において，後頭部と左肩に近い位置にないはずの背中と誤検出されるなど，単に距離の近いことによる誤検出以外も起きていることが判明した．

表 1 実験 1 (立位)：左手でタップした部位の推定精度 (単位：%)

部位	A	B	C	D	E	平均
後頭部	100	100	100	90	100	98
右肩	100	80	70	100	100	90
背中 (左側)	60	50	100	100	20	66
左腰	40	50	0	0	100	38
タップ検出率	100	100	100	100	100	100

表 2 実験 1 (立位)：右手でタップした部位の推定精度 (単位：%)

部位	A	B	C	D	E	平均
後頭部	70	100	100	90	100	92
左肩	90	100	70	100	100	92
背中 (右側)	20	30	100	100	100	70
右腰	100	90	0	80	70	68
タップ検出率	97.5	100	100	97.5	100	99

表 3 実験 1 (立位)：左手のタップで検出された部位の混同行列

検出部位 \ 正解部位	後頭部	右肩	背中	左腰
後頭部	49	0	1	0
右肩	5	45	0	0
背中	0	0	33	17
左腰	0	0	31	19

表 4 実験 1 (立位)：右手のタップで検出された部位の混同行列

検出部位 \ 正解部位	後頭部	左肩	背中	右腰
後頭部	46	0	2	0
左肩	3	46	1	0
背中	0	0	35	15
右腰	0	0	16	34

6.2 実験 2 について

実験 2 を行った際の部位ごとの検出精度およびタップ検出精度を表 5, 6 に示す。また、混同行列を表 7, 8 に示す。実験 2 全体を通してのタップ検出精度はタップ検出精度は平均 100%, 部位推定精度は平均 78.25% であった。実験 2 でも、腰の推定精度については低い結果となり、個人差の大きい結果であることは変わらなかった。しかし、実験 1 では一切腰が検出されなかった参加者でも検出されるなど、変化がみられた。また、全体としては、実験 1 と比較すると、肩の検出精度が低下し、背中中の検出精度が上昇する傾向がみられた。次に、混同行列に着目すると、背中と腰を互いに誤検出する傾向については変わらないが、左手において背中と後頭部を混同しているのが 1 回、右手において後頭部や肩と背中を混同しているのが 7 回あり、隣接部位ではない部位での誤検出が実験 1 よりも多くみられた。これについては、身体後方ででのトラッキングの値が精度の不足により異常値を出力した、内蔵カメラのトラッキングにおいて、右手がトラッキングロストした直後に一部写った左手と右手を誤認識されたなどの原因が考えられる。

6.3 アンケート結果

手首にデバイスを装着することによる手の動かしにくさについては、5 人中 3 人があったと回答した。特に、Joy-Con の ZL, ZR ボタンが手の甲に当たって痛いという意見や、Joy-Con が縦に装着されていることから手首が曲げづらいという意見が得られた。

次に、実験を通してタップを続けた結果痛くなった場所については、手が痛くなった、背中を叩く際に肩が痛くなったなどのコメントが得られた。

実験 1, 2 の中で、タップがしにくい部位については、立位、座位両方の場合において、実験参加者全員が、背中が叩きにくいという回答をした。他の部位に関しては、「お尻に関しては座ってる分叩く面積が変わってしまってどこを叩くべきかわからない」、「座った時にお尻が叩きづらく

て横を叩いていた」(注：本稿では臀部を含めた腰回りの大きな部位を腰という表記で統一している) という意見があったが、これについては、座った際にタップ可能な面積が減少することや、体格によっては推定精度が低くなり、判定が厳しくなったことが原因の一つだと考えられる。逆に、タップしやすい部位については、(特に立位について) 5 人中 3 人が「お尻を叩くのは楽しい」、「お尻を叩くのは気分がいい」、「お尻を叩いてもいたくないので楽しい」など、肯定的な意見が得られた。

タップした部位に割り当てたい操作については、実験参加者の多くが、通常の操作ではワンアクションでできない操作や、メニュー呼び出し、スタートボタン、決定ボタンなどの決まった操作を割り当てることで、ワンタップで機能呼び出したいという意見が多かった。また、アプリケーションに関しては、4 択の選択肢を選ぶ操作の多いアプリケーションでの利用や、ハンドトラッキングで操作をすることの多い AR での利用を行いたいという意見があった。

また、全体的に気になった点については、(Quest Link 用の) コードが邪魔になった、HMD をかぶっている状態で後頭部を叩くと HMD (のエリートストラップ) に当たって痛いなどの点が挙げられていた。

6.4 トラッキング精度

手首のトラッキング精度について検証を行うため、学習用データセットを作成した時と同じ環境で検証用データを作成し、実験時に用いた学習済みモデルに対してオフラインで精度を測定した。検証用データセットは、カメラの視野外に出る場合と視野内に戻る場合を繰り返して作成している (学習用データセットに関しても同様)。なお、位置の推論結果に関しては、実験時と同様の条件にするため、視界に入った最後の手首の位置に推論結果の変化量を加えた値を実際の推論結果として誤差を測定した。検証用データ全てのサンプルの位置、回転に関する誤差を図 4 に示す。その結果、位置に関する中央値誤差は 6.95 cm, 平均絶対

表 5 実験 2 (座位)：左手でタップした部位の推定精度 (単位：%)

部位	A	B	C	D	E	平均
後頭部	100	100	100	100	100	100
右肩	90	70	80	80	100	84
背中 (左側)	90	90	90	90	40	80
左腰	70	60	0	70	30	46
タップ検出率	100	100	100	100	100	100

表 7 実験 2 (座位)：左手のタップで検出された部位の混同行列

正解部位 \ 検出部位	検出部位			
	後頭部	右肩	背中	左腰
後頭部	50	0	0	0
右肩	7	42	1	0
背中	1	0	40	9
左腰	0	0	27	23

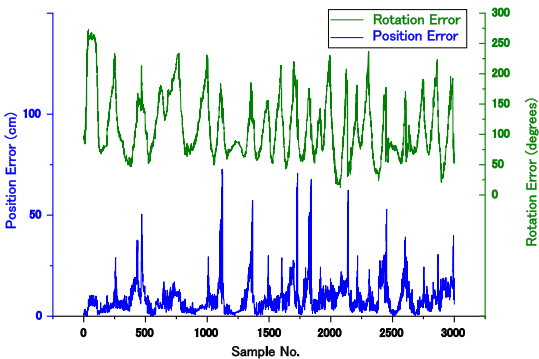


図 4 検証用データの全サンプルにおける位置と回転の誤差

値誤差は 9.03 cm, 最大誤差は 72.80 cm となった. 結果を見ると, 中央値誤差や平均絶対値誤差で見ると誤差は大きくはないものの, 全サンプルのグラフでは, 瞬間的に大きな誤差になっているサンプルが複数箇所あり, 誤差が十分実用的な範囲に収まっているとは言えないことがわかる.

また, 出力された手首の回転の誤差については, 中央値誤差は 95.86°, 平均絶対値誤差は 111.2°, 最大誤差は 273.3° となった. 回転の誤差は, 中央値誤差, 平均絶対値誤差, グラフの全てにおいて誤差が大きく, グラフの値が 0 に近い点が少ないことから, 定数分の誤差が発生している可能性がある. また, 定数分の誤差を補正したとしてもサンプル間の誤差の変化が非常に大きいため, 機械学習モデルの学習自体がデータの規模や品質の不足により収束していない可能性がある.

7. 議論

本章では, 本手法に関する課題や将来的な拡張について議論を行う.

7.1 ユーザの姿勢変化や体格による影響

本研究の実装では, 身長の高さに関する分布と身体を構

表 6 実験 2 (座位)：左手でタップした部位の推定精度 (単位：%)

部位	A	B	C	D	E	平均
後頭部	60	100	100	100	100	92
左肩	70	50	70	100	100	78
背中 (右側)	80	100	100	100	80	92
右腰	70	90	20	50	40	54
タップ検出率	100	100	100	100	100	100

表 8 実験 2 (座位)：右手のタップで検出された部位の混同行列

正解部位 \ 検出部位	検出部位			
	後頭部	左肩	背中	右腰
後頭部	46	0	4	0
左肩	8	39	3	0
背中	0	0	46	4
右腰	0	0	23	27

成する部位の大きさの分布が正規分布に従う仮定の下で, HMD の座標から身長を算出し, その身長を元に相対的な部位の位置を統計的数値から決定している. したがって, 実験参加者が大きく背中を曲げた状態であったことや, 一部の実験参加者の体格がその身長での平均的体格と大きく異なっていたために, 正確な検出ができなかった可能性がある. 具体的には, 胴や足の長さの比率が統計データの分布から算出した値と離れていた可能性が考えられる. この問題については, 頭と両手首の 3 点トラッキングを元にした機械学習によるボディトラッキングや, 頭の姿勢情報による身体の傾きの補正, 推定に距離だけでなく手首の姿勢情報を用いることなどで改善する可能性がある.

7.2 ハンドトラッキングの精度の影響

今回の実装では, 先行研究 [9] と同様の機械学習モデルを用いた. しかし, 検証用データセットでの数値は中央値誤差と平均絶対値誤差については問題なかったが, 誤差が極端に大きな点があった. この点については, 身体後方で学習データを追加で集めることで改善すると考えられる. また, 今回のようなハンドトラッキング手法を用いる場合, 精度を上げたとしても IMU のノイズなどが原因で数 cm から 10 数 cm 程度の誤差が発生する. このような誤差により, 腰, 背中, 肩などの大まかな部位の検出は行うことができるが, 1 つの部位を複数の領域に分割するような細かな検出を行うことは非常に難しいと予想される. しかし, 全身を入力領域として使った場合, 大まかな分割だとしても多くのパターン数を用意できると考えられるため, 大まかな部位がほぼ確実に検出できる程度の精度まで改善すれば十分だと考えられる.

7.3 VR 以外への適用

本研究では MR での入力領域, 入力語彙の拡張を目指したものの, 手法そのものはユーザの身体部位 (または頭部) と手をトラッキングできれば利用可能である. 今回は

Meta Quest 2での実装を行ったが、Hololens 2などのARデバイスや、カメラによるトラッキングなどとの組み合わせでも本手法が適用できると考えられる。また、今回は手首に装着するデバイスとして Joy-Con を利用したが、スマートウォッチに搭載された IMU を利用することでも実現できると考えられる。したがって、AR などの手を使って操作をするような場面や、スマートウォッチなどを装着して簡単な操作を行う場面に対しても本手法の適用範囲を広げることが期待できる。

7.4 操作のデザインによる拡張性

実験では、手が内蔵カメラの視界の範囲内に入ってからタスクを行うようにすることで、手がトラッキング範囲外にある時間を短縮した。今回は単純にタップした部位の推定までを実装したが、実際に提案手法を利用する際には、通常のハンドトラッキングによるジェスチャと併用して組み込まれることが想定される。したがって、どのように手法を組み合わせるかを工夫してデザインすることで、実験時と同様に手がトラッキング範囲外にある時間を短縮し、精度の低下を抑制できると考えられる。また、単純なタップ操作だけでなく、片手でタップ操作、もう片手でジェスチャ操作を行うなどの複数操作の組み合わせや、各部位に対する最適な操作の割り当てなどの操作のデザインを行うことで、単純な検出部位の数にとどまらないパターンの入力が行えることが期待される。

8. まとめと展望

本稿では、自分の身体背面部をタップすることで VR/MR 空間上での入力領域を拡張するシステムを提案した。また、提案したシステムに対して実験を行い、タップ検出精度とタップした部位の推定精度について評価した。立位姿勢で指定した4つの部位をタップする実験では、平均76.25%、座った状態でタップする実験では平均78.25%の部位検出精度が得られた。また、実験結果や想定される課題に関して、今後の改善に向けた議論を行った。今後は、ハンドトラッキング精度の改善や、姿勢変化の影響を軽減するための追加実装により、タップした部位の推定精度の向上や、タップ可能な部位数の増加を目指す。

参考文献

[1] Bergström, J. and Hornbæk, K.: Human-Computer Interaction on the Skin, *ACM Comput. Surv.*, Vol. 52, No. 4 (2019).

[2] Cao, Z., Simon, T., Wei, S.-E. and Sheikh, Y.: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields, *CVPR'17* (2017).

[3] Caserman, P., Garcia-Agundez, A., Konrad, R., Göbel, S. and Steinmetz, R.: Real-time body tracking in virtual reality using a Vive tracker, *Virtual Reality*, Vol. 23, pp. 155-168 (2019).

[4] Jiang, J., Strelci, P., Qiu, H., Fender, A., Laich, L., Snape, P. and Holz, C.: AvatarPoser: Articulated Full-Body Pose Tracking from Sparse Motion Sensing, *Proceedings of European Conference on Computer Vision*, Springer (2022).

[5] Winkler, A., Won, J. and Ye, Y.: QuestSim: Human Motion Tracking from Sparse Sensors with Simulated Avatars, *SIGGRAPH Asia 2022 Conference Papers*, SA '22 (2022).

[6] Chiu, P., Takashima, K., Fujita, K. and Kitamura, Y.: Pursuit Sensing: Extending Hand Tracking Space in Mobile VR Applications, *Symposium on Spatial User Interaction*, SUI '19 (2019).

[7] Shen, S., Wang, H. and Roy Choudhury, R.: I Am a Smartwatch and I Can Track My User's Arm, *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '16, p. 85-96 (2016).

[8] Wei, W., Kurita, K., Kuang, J. and Gao, A.: Real-Time 3D Arm Motion Tracking Using the 6-axis IMU Sensor of a Smartwatch, *2021 IEEE 17th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, pp. 1-4 (2021).

[9] Strelci, P., Armani, R., Cheng, Y. F. and Holz, C.: HOOV: Hand Out-Of-View Tracking for Proprioceptive Interaction Using Inertial Sensing, *CHI '23* (2023).

[10] Harrison, C., Ramamurthy, S. and Hudson, S. E.: On-body interaction: armed and dangerous, *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction*, TEI '12, p. 69-76 (2012).

[11] Gustafson, S. G., Rabe, B. and Baudisch, P. M.: Understanding palm-based imaginary interfaces: the role of visual and tactile cues when browsing, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, p. 889-898 (2013).

[12] Harrison, C., Tan, D. and Morris, D.: Skinput: Appropriating the Body as an Input Surface, *CHI '10*, p. 453-462 (2010).

[13] Mi, T.-W., Wang, J.-J. and Chan, L.: LapTouch: Using the Lap for Seated Touch Interaction with HMDs, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, Vol. 7, No. 3 (2023).

[14] Cheng, W.-W. and Chan, L.: BodyTouch: Investigating Eye-Free, On-Body and Near-Body Touch Interactions with HMDs, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, Vol. 7, No. 4 (2024).

[15] He, K., Zhang, X., Ren, S. and Sun, J.: Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification, *ICCV'15* (2015).

[16] Kingma, D. P. and Ba, J.: Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980* (2014).

[17] Hartikainen, J., Solin, A. and Särkkä, S.: Optimal filtering with Kalman filters and smoothers, *Department of biomedical engineering and computational sciences, Aalto University School of Science, 16th August* (2011).

[18] Zhou, Y., Barnes, C., Lu, J., Yang, J. and Li, H.: On the Continuity of Rotation Representations in Neural Networks, *CVPR'19* (2019).

[19] 中村邦子, 武本歩未, 大塚美智子: 2014-2016 日本人の人体計測データの分析結果, *繊維製品消費科学*, Vol. 63, No. 5, pp. 313-322 (2022).