

# MediaPipeを用いたジェスチャ認識による メタバース空間での直感的操作手法

二村 拓<sup>1,a)</sup> 藤枝 元輝<sup>†1,b)</sup> 岩井 将行<sup>1,†1,c)</sup>

**概要:** IT 技術の発展により、日常生活で IT 機器を利用する機会は年齢を問わず増えている。しかし、キーボードやマウスなど従来の入力手段は習熟が必要で、IT 操作に不慣れな利用者には負担となる場合がある。この課題に対し、手や指の動きを直接入力として用いるハンドジェスチャ操作は、身体の動きと操作結果の対応が直感的で、事前学習の負担を軽減できる入力手法として注目されている。一方で、既存手法の多くは画面接触や専用デバイスを必要とするため、使用環境の制約や装着の負担、導入コストといった課題がある。

本研究では、汎用的な Web カメラ映像を用い、MediaPipe Hands によって手指の動きを高精度に推定し、ハンドジェスチャで IT 機器を操作する手法を提案する。

検証には、教育分野での利用実績がある高拡張性のサンドボックス型 3D ゲーム「Minecraft Java Edition」を採用し、先行手法との比較実験と操作性アンケートを実施した。その結果、回答者の 93.3% が本手法をより直感的と評価した。以上の結果から、提案手法は操作性の向上に寄与することが示された。

## 1. はじめに

IT 技術の発展により、日常生活で IT 機器を利用する機会は年齢に関係なく増えている。しかし、キーボードやマウスなど従来の入力手段は習熟が必要で、IT 操作に不慣れな利用者には心理的負担となる場合がある。この課題を解決する方法として、手や指の動きを直接入力とするジェスチャ操作が注目されている。身体動作と操作結果の対応が直感的であるため、操作理解が容易で、利用者の負担軽減につながると考えられる。

藤枝ら [1] は、ハンドジェスチャで Minecraft のキャラクターを操作するシステムを開発した。Minecraft は教育分野での利用実績があり、高拡張性のサンドボックス型 3D ゲームとして実験環境に適している。同研究では、動的ジェスチャ取得用に M5Stack 用ジェスチャユニット、静的ジェスチャ取得用に M5Stack CoreS3 のカメラモジュールを使用し、取得データは Node-RED サーバや画像推論サーバを経由して Minecraft サーバに送信される構成である。Spigot サーバを用いることで、Java プラグインによる柔軟なゲーム内制御が可能となっている。



図 1 先行研究で利用したデバイス

しかし、この手法は専用デバイス依存であることや、認識精度の制約、ジェスチャと操作結果の対応が分かりにくく直感性に欠ける点が課題である。さらに、デバイス装着や準備の手間が導入や継続利用の障壁となる可能性も指摘されている。

表 1 先行研究で使用しているハードウェア

名称	用途
M5Stack CoreS3	映像取得, データ送信
M5Stack 用ジェスチャユニット	動的ジェスチャの取得

<sup>1</sup> 東京電機大学未来科学部情報メディア学科

<sup>†1</sup> 現在, 東京電機大学大学院未来科学研究科情報メディア学専攻

a) taku@cps.im.dendai.ac.jp

b) genki@cps.im.dendai.ac.jp

c) iwai@cps.im.dendai.ac.jp

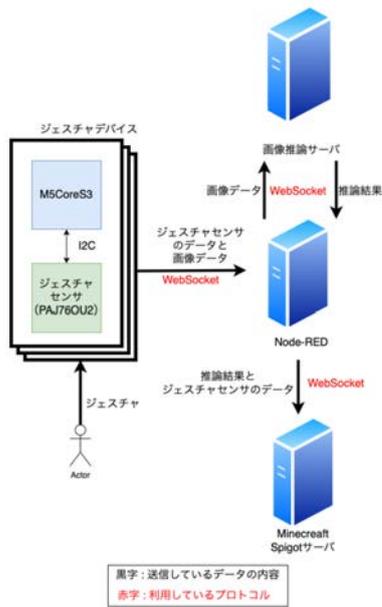


図 2 先行手法のシステム構成図



図 3 図 1 のデバイスを操作している様子

## 2. 関連研究

### 2.1 ハンドジェスチャを用いた操作手法

Weon ら [2] は、ハンドジェスチャを入力手段としたドローン操作手法を提案している。本研究は、自律走行車に代わる移動手段としてドローンを安全に制御するための手法である。仮想環境上での検証により、9 種類のジェスチャによる操作が可能であることを示した。

Pame ら [3] は、ハンドジェスチャを用いたマウスカーソル操作に関する研究を行った。指の状態を認識することで、カーソル移動やクリック動作を実現している。また、手の移動量をカーソルの挙動に対応させる独自のマッピング方式を導入することで、操作の直感性向上を図っている。

芝田ら [4] は、前腕部から取得した表面筋電位を用いてジェスチャを判別し、個人認証へ応用する手法を検討した。同一のジェスチャであっても個人ごとに筋活動が異なる点に着目し、被験者 5 名に筋電計 (Myo) を装着させ、6 種類のジェスチャを実施させてデータを収集した。その結果、平均識別率は 70.7 % であったが、精度向上に向けた課題も残されていることが示された。

### 2.2 パソコン操作における課題意識

広瀬ら [5] は、パソコン操作に対する心理的抵抗感が課題遂行に及ぼす影響について検討した。その結果、苦手意識を有する被験者では、Excel を用いた作業課題の達成率が相対的に低下する傾向が示された。

### 2.3 Minecraft の教育活用

Vořtinár ら [6] は、Minecraft 教育版を用いたプログラミング学習の効果について検証を行った。13 歳から 14 歳の学生を対象とした実験の結果、参加者全員が学習の継続に前向きな意向を示しており、教育分野における Minecraft 活用の有効性が示唆された。

岩橋ら [7] は、宮城教育大学附属中学校を対象に、Minecraft 上で仙台市上杉地区の地層構造を再現し、地層および地質時代に関する理解を促す学習活動を実施した。具体的には、作成したクイズに生徒が取り組む形式を採用し、地学分野への関心喚起を図った。約 160 名の生徒を対象としたアンケート調査では、興味に関する評価項目の平均値が満点である 7 点に近い結果となり、Minecraft を用いた教材が学習意欲の向上および理解促進に寄与することが確認された。

### 2.4 Minecraft の教育以外での活用例

ウラタら [8] は、Minecraft を基盤としたデジタルツイン型プラットフォームである EnvBridge を提案している。EnvBridge は、温度、湿度、黒球温度および風速といった環境データを入力として熱中症リスクを算出し、その結果を視覚的に提示することで予防への活用を目的としている。Minecraft 上での可視化を採用することにより、親しみやすい操作性と高い拡張性を維持しつつ、導入時のコストや運用面での負担を抑制できる点が特徴である。

### 2.5 MediaPipe を用いた手指トラッキング技術

Zhang ら [9] は、単眼 RGB カメラ映像から手指をリアルタイムに追跡する MediaPipe Hands を提案している。手のひら検出と手指ランドマーク推定の二段階構成により、21 点の手指ランドマークを高精度に推定可能で、専用センサ不要かつモバイル端末でも低遅延で動作する。

### 2.6 MediaPipe を用いたハンドジェスチャ操作

Jo and Kim [10] は、MediaPipe Hands を用いたモバイル VR 向けハンドジェスチャ分類手法を提案している。本研究では、単眼カメラ映像から推定される 21 点の手指ランドマークを用い、指関節の角度情報に基づいて拳状ジェスチャを認識し、Unity 上で実装したゲーム操作に応用している。また、MediaPipe の複数実装 (C++ ポート版、TensorFlow Lite 版、Unity Barracuda 版) について、PC およびモバイル環境での処理速度と認識性能を比較評価しており、実行環境に応じた実装選択の重要性を示している。

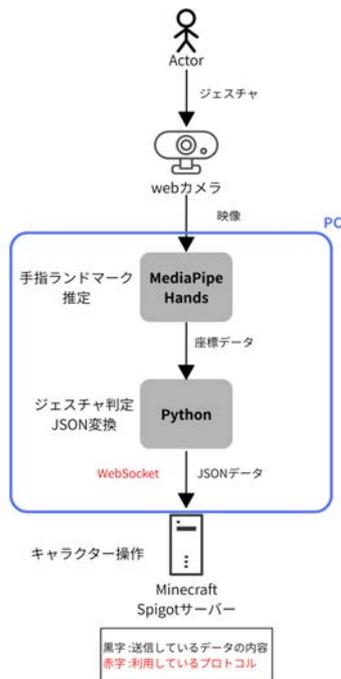


図 4 提案手法システム構成図

本研究は、MediaPipe を用いたジェスチャ操作が、専用デバイスを用いずに直感的な仮想環境操作を実現できる可能性を示した点で意義がある。

### 2.7 関連研究のまとめ

以上の関連研究から、ハンドジェスチャを用いた直感的な操作手法は、IT 機器に対して苦手意識を有する利用者に対し、操作習得に要する時間や心理的負担を低減し、容易な利用を可能とすると考えられる。また、Minecraft は教育分野およびエンターテインメント分野の双方において高い認知度を持ち、実験環境としての適性も高い。

本研究では MediaPipe を用いた高精度なハンドジェスチャ認識技術を活用し、直感的に Minecraft のキャラクターを操作するアプローチを提案する。これにより、従来のキーボードやマウス操作に苦手意識を持つ利用者に対して、パソコン操作に伴う心理的負担を軽減し、学習意欲の向上が期待される。

## 3. システム構成

図 4 にシステム構成図を示す。本システムは、Web カメラで取得した映像を MediaPipe で解析するジェスチャ認識部と、認識結果を受信して Minecraft の操作に反映する Spigot サーバから構成される。両者は WebSocket により接続され、リアルタイムに連携する。

### 3.1 ジェスチャ取得システム

本研究では、Web カメラを用いて手指動作を取得するジェスチャ入力システムを構築した。取得した映像は Me-

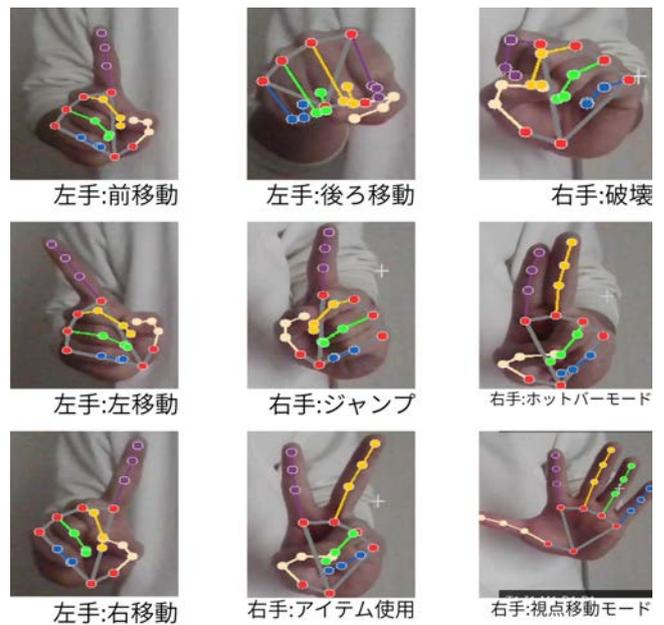


図 5 判定するジェスチャー一覧

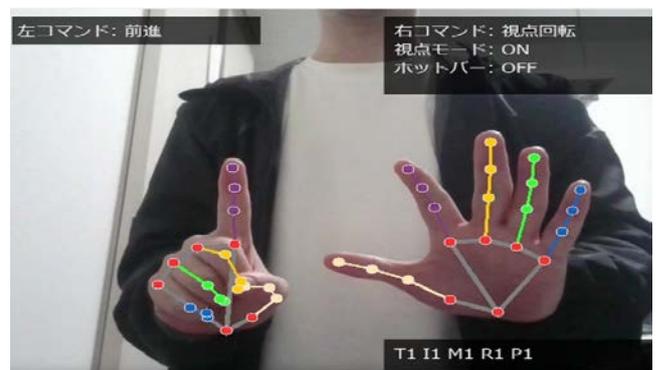


図 6 提案手法のシステムを操作している様子

diaPipe Hands に入力され、21 点の手指ランドマークがリアルタイムに推定される。これらの座標情報から、指の開閉状態や関節角度などの特徴量を算出し、ジェスチャを識別する。本システムでは、専用デバイスを使用せず、一般的な Web カメラのみで動作する構成とすることで、導入コストの低減と設置環境への依存を抑える点を重視した。

### 3.2 ジェスチャ判定システム

ランドマーク座標に基づき、指の伸展度、手首方向、手の向きなど複数の特徴量を用いてジェスチャを判定する方式を採用した。本研究では、機械学習による分類ではなく、閾値に基づくルールベースの判定手法を用いることで、処理負荷の軽減と判定基準の可視化を図った。さらに、動作の揺れや一時的な誤判定を防ぐため、単一フレームではなく、複数フレームの判定結果が一定条件を満たした場合にのみジェスチャを確定する連続判定処理を導入した。これにより、誤認識を抑制し、安定した操作を実現した。

### 3.3 データ送信システム

認識したジェスチャ情報は JSON 形式に変換され、WebSocket を介して Minecraft 側の Spigot サーバに送信される。WebSocket を用いることで、Python 側の認識処理と Minecraft 内の操作処理をリアルタイムに連携させることが可能である。

### 3.4 プラグインによるジェスチャ操作

表 2, 表 3 に本研究で使用可能なジェスチャと、それに対応する Minecraft 内の動作を示す。本システムでは、Minecraft 用のサーバ環境上に独自に開発したプラグインを適用し、認識されたジェスチャをゲーム内操作として反映している。ジェスチャの設計にあたっては、キャラクターの移動や視点操作が直感的に対応するよう配慮した。これにより、操作方法を事前に詳細に学習しなくても、利用者が自然な動作でキャラクターを操作できることを意図している。また、プラグインは改変が可能であるため、ジェスチャと対応する動作を柔軟に追加・変更でき、利用目的やユーザ特性に応じた操作体系の拡張が可能である。

表 2 左手ジェスチャと対応動作一覧

ジェスチャ	対応動作
人差し指が上を指す	前に移動する
人差し指が上から 30 度以上 90 度以下で左に傾ける	左に移動する
人差し指が上から 30 度以上 90 度以下で右に傾ける	右に移動する
人差し指が上から 90 度以上傾いている	後ろに移動する
指をすべて閉じてグーの形をしている	後ろに移動する

## 4. 評価

ここでは、アンケート内容とその結果を示す。アンケート調査は本学の協力者などを対象に実施し、計 15 名から回答を得た。

### 4.1 アンケート結果

以下に、提案手法に関するアンケート結果を示す。

- (1) 提案手法の認識精度はよかったですか (1: 最小, 5: 最大)  
回答 (1):0.0%, (2):0.0%, (3):13.3%, (4):66.7%, (5):20.0%
- (2) 提案手法の視点の移動角度は適切でしたか (1: 最小, 5: 最大)  
回答 (1):0.0%, (2):13.3%, (3):20.0%, (4):53.3%, (5):13.3%
- (3) 2つの手法のうちどれが最も直感的に利用できましたか

表 3 右手ジェスチャと対応動作一覧

ジェスチャ	対応動作
拳状 (全指屈曲) で上方向にスワイプ	ブロック破壊
人差し指を伸ばした状態で上方向にスワイプ	ジャンプ
人差し指と中指を伸ばし、二指を揃えて上方を指す	ホットバーモードに移行
ホットバーモード中に二指を左へ傾ける	ホットバーのカーソルが左に移動
ホットバーモード中に二指を右へ傾ける	ホットバーのカーソルが右に移動
人差し指と中指を伸ばし、二指を揃えて上方を指す状態を解除	ホットバーモード終了
人差し指と中指を伸ばし、開いた状態で上方向にスワイプ	ブロック設置
全指伸展 (パー) 状態で 0.5 秒間保持	視点移動モードに移行
視点移動モード中に手を移動	移動した方向に視点が移動
全指伸展状態を解除	視点移動モード終了

回答 提案手法:93.3%, 先行手法:6.7%

- (4) その操作手法を「より直感的」と感じた理由は何ですか
  - (a) 提案手法
    - ・精度が高かったから (5名)
    - ・思い通りに動いたから (1名)
    - ・先行手法よりも動かしやすかったから (3名)
    - ・移動がスムーズだった (2名)
    - ・画面上にどの操作が実行されているか分かるから (1名)
    - ・ジェスチャと行動がリンクしていた (2名)
  - (b) 先行手法
    - ・視点移動がシンプルだった (1名)
- (5) ジェスチャ操作でゲームキャラクターを操作するのは楽しかったですか (1: 最小, 5: 最大)  
回答 (1):0.0%, (2):0.0%, (3):26.7%, (4):73.3%, (5):0.0%
- (6) より快適に操作するために必要なことや改善点はありますか
  - ・視点移動をより滑らかにする (5名)
  - ・精度を向上させ、誤作動を減らす (3名)
  - ・手が疲れないように工夫する (3名)
  - ・機能を増やす (3名)
- (7) キーボード、先行手法、提案手法のうちどれを使いたいと思いますか  
回答 キーボード:86.7%, 先行手法:0.0%, 提案手法:13.3%

(8) その理由を教えてください

(a) キーボード

- ・キーボードに慣れているから (7名)
- ・疲れにくいから (2名)
- ・操作しやすいから (2名)
- ・ジェスチャではキャラクタが思い通りに動かせないから (1名)

(b) 提案手法

- ・初心者ならジェスチャ操作のほうが使いやすいと思う (2名)

(c) 先行手法

- ・回答なし

## 4.2 考察

アンケート結果では、回答者の 93.3% が先行手法より提案手法の方が直感的に操作できると回答した。これは、MediaPipe Hands による認識精度の向上や、画面上で実行中のジェスチャを可視化する機能により、利用者が自身の動作と Minecraft 上の動作結果との対応を即座に把握できたためと考えられる。これにより、操作意図とシステム動作の乖離が低減され、直感的操作感が向上した。

一方、キーボード操作との比較では、回答者の 86.7% がキーボード操作を好むと回答した。その理由として、キーボードの方が入力が速い点が挙げられるが、慣れていない利用者にとっては提案手法の方が使いやすい可能性も示された。

以上より、本手法は従来入力手段の代替ではなく、IT 操作に不慣れた利用者や新しい操作体験を求める場面で有効な入力手段となる可能性がある。

## 5. まとめと展望

本研究では、汎用 Web カメラと MediaPipe Hands を用いて手指動作を認識し、WebSocket 経由で Minecraft を操作するハンドジェスチャ手法を構築した。専用デバイス不要で導入負担を抑えつつ直感的操作を実現し、アンケートでは 93.3% が先行手法より直感的と評価した。

今後は、誤認識や長時間利用時の疲労軽減のため認識精度向上やジェスチャ設計の改善、利用者による動作カスタマイズ、視線や音声など他入力手法とのマルチモーダル化を検討する。さらに Minecraft 以外の仮想環境への適用を通じ、本手法の汎用性と実用性を評価していく予定である。

## 6. 謝辞

本研究を進めるにあたり、多くの方々にご支援、ご助言をいただきました。日頃より有益なご意見をいただいた研究室の皆様に感謝いたします。

## 参考文献

- [1] 藤枝元輝, 岩井将行, "IoT ノードと機械学習サーバを用いた指ポーズ推定によるメタバース空間での操作手法", "情報処理学会ヒューマンコンピュータインタラクション研究会", PaperID=AA1221543X, 2025 年 1 月 15 日発表
- [2] Juyeon Weon, Taein Yong, and Jaeho Kim. Camera-based virtual drone control system using two-handed gestures. In 2024 IEEE International Conference on Meta-verse Computing, Networking, and Applications (Meta-Com), pp. 291–296, 2024.
- [3] Yash Gajanan Pame and Vinayak G Kottawar. A novel approach to improve user experience of mouse control using cnn based hand gesture recognition. In 2023 7th international Conference On Computing, Communication, Control And Automation (ICCUBE), pp. 1–6, 2023.
- [4] 芝田龍正, 三上剛, 秋月拓磨, 大前佑斗, 高橋弘毅. 個人認証を念頭に置いた表面筋電位を用いたジェスチャ識別. 知能と情報, Vol. 33, No. 1, pp. 549–554, 2021.
- [5] 広瀬啓雄, 難波和明. パソコン操作において苦手意識が課題達成に与える影響の分析. 経営情報学会 全国研究発表大会要旨集, Vol. 2002f, pp. 12–12, 2002.
- [6] P. Voštinár and R. Dobrota. Minecraft as a tool for teaching online programming. In 2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO), pp. 648–653, 2022.
- [7] 岩橋純子, 安藤明伸, 西岡芳晴, 川畑大作, 白石喬久. コンピュータゲーム「minecraft」を用いた仙台市上杉地区の地質に関する教材開発. 地学雑誌, Vol.131, No. 4, pp. 463–472, 2022.
- [8] 英寿ウラタ, 伶馬山崙, 昭徳丁, 渡邊, 岩井将行. 4 方位のリアルタイム簡易風速センシングによる体感温度予測及び簡易熱中症リスク啓蒙デジタルツインシステム. 第 86 回全国大会講演論文集, 第 2024 巻, pp. 341–342, mar 2024.
- [9] Zhang, F., Bazarevsky, V., Vakunov, A., Tkachenka, A., Sung, G., Chang, C.-L., & Grundmann, M., "MediaPipe Hands: On-device Real-time Hand Tracking," *arXiv preprint*, arXiv:2006.10214, 2020.
- [10] B. J. Jo and S. K. Kim, "A Study on the Classification of Hand Gesture for Mobile Virtual Reality with MediaPipe," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 21s, pp. 3555–3559, 2024. [Online]. Available: