

# HoloLens2 を用いた自由な全身のジェスチャを定義可能な 機器操作システムの提案

○望月 勇伸（中央大学），Sarthak PATHAK（芝浦工業大学），梅田 和昇（中央大学）

## Proposal of a Device Operation System Enabling User-Defined Full-Body Gestures with HoloLens2

○ Yushin MOCHIZUKI (Chuo University), Sarthak Pathak (Shibaura Institute of Technology),  
and Kazunori UMEDA (Chuo University)

**Abstract:** This paper proposes a system that enables device operation through user-defined full-body gestures using HoloLens2. The system first acquires the user's posture through external cameras and automatically reflects it on a virtual avatar in HoloLens2. The user then manipulates the avatar directly to visually define gesture postures and tolerance ranges for recognition. After the definition process, gesture recognition and device operation can be performed solely with external cameras, without wearing HoloLens2. An evaluation experiment was conducted to verify the operational accuracy of the proposed method. Furthermore, we clarified the effects of reducing workload through user-defined gestures and assessed the overall usability of the system.

### 1. 序論

近年，HCI (Human Computer Interaction) の多様化が進み，扱うデバイスや通信機器の増加に伴い，従来の手法に加えて各環境に適した操作手法の導入が求められている。中でもジェスチャ認識は，利便性の高い家電操作や非接触インターフェースとして注目されており，作業支援や医療など幅広い分野での活用が期待されている。しかし，利用環境や個人差により，特定のジェスチャが実行しづらい場合があるため，パーソナライズ化されたジェスチャ認識システムの実現が重要となる。

我々はこれまでに，ユーザが右手の任意の姿勢を定義可能なジェスチャ操作システムを提案し，HoloLens2を用いたアバター表示により基礎的な有効性を確認した<sup>1)</sup>。しかし，本システムは右腕に限定されており，全身動作を用いた自然かつ多様な操作には対応できないという課題があった。本稿はその拡張版として，全身の任意ジェスチャを定義可能な機器操作システムを提案する。

本システムでは，HoloLens2 上に自由な姿勢を取るアバターを表示し，直感的に全身でのジェスチャ定義を可能にする。さらに，本システムは，外部カメラとHoloLens2 を連携させ，定義したジェスチャ情報を送受信することで，定義時の負担軽減や定義後の HoloLens2 を用いない操作を実現する。これにより，非接触で任意の全身ジェスチャを定義可能とし，スマートホーム，医療，産業分野など多様な応用を見据えている。また，被験者実験を実施し，NASA-TLX や SUS を用いた評価を通じて，操作精度，ユーザビリティ，作業負荷の観点から定量的に検証を行う。

### 2. 関連研究

ユーザビリティの高いシステムを実現するため，ジェスチャ操作をパーソナライズ化する研究が行われている。Few-Shot 学習を応用することで，ユーザごとに適応したハンドジェスチャ認識を可能にするシステムが提案されている<sup>2)3)</sup>。しかし，これらは手の動作に限定されており，定義可能なジェスチャ数や表現力に



Fig. 1 System Overview

制約があるという課題が残されている。

また，AR (Augmented Reality) や MR (Mixed Reality) を用いたジェスチャ定義の研究も進められている。Schäfer らは，任意の片手ジェスチャを少しき回のデモンストレーションで登録可能な AnyGesture を提案した<sup>4)</sup>。さらに，Shen らは自然で連続的な動作の中から特定のキージェスチャを正確に検出する Gesture Spotter を開発した<sup>5)</sup>。これらのアプローチは，没入型環境における柔軟なインタラクションを実現する点で有効である。しかし，操作対象は VR/MR 環境内の仮想オブジェクトや UI に限定されており，現実世界の物理デバイスなど多様な対象へは応用が困難である。

そこで本研究では，ユーザが任意に定義可能な全身ジェスチャを用いて現実世界の機器操作を実現するシステムを提案する。

### 3. 提案手法

#### 3.1 概要

本研究で提案するシステムは，アバター姿勢の自動調整，HoloLens2 を用いたジェスチャ定義，外部カメラによるジェスチャ認識の順で実行する。これにより，ユーザが任意に定義した全身ジェスチャを用いて機器操作を直感的かつ柔軟に行うことを可能とする。

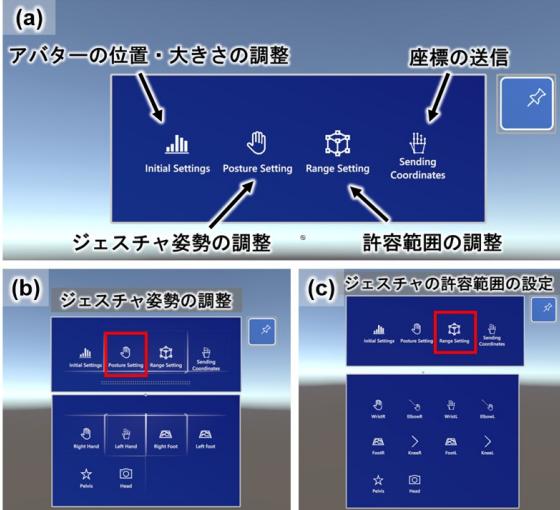


Fig. 2 The Role of Each Button

### 3.1.1 システム構成

本システムでは、Fig.1 に示すように、ユーザは HoloLens2 を装着し、三人称視点で等身大アバター<sup>6)</sup>を確認しながらジェスチャを定義する。この際、アバターの姿勢を初期状態から手作業で構築することは、ユーザに大きな負担を与える。そこで、外部カメラによりユーザの姿勢を取得し、その情報を HoloLens2 と送受信することで、アバターの姿勢を自動的に調整する。

自動調整後、ユーザはアバターの姿勢を微調整し、操作のばらつきや自由度を考慮した許容範囲を設定することでジェスチャを定義する。定義が完了すると、情報は外部カメラによるジェスチャ認識モジュールへ送信され、定義済みデータに基づいて認識が行われる。これにより、ユーザは HoloLens2 を装着せずに操作を実行可能となる。

### 3.1.2 システム環境

ジェスチャ定義は HoloLens2 上に表示されるアバターを介して行い、ユーザが視覚的に確認しながら調整を進めることができる。一方で、アバターの初期姿勢生成およびジェスチャの認識処理は、外部に設置したカメラを用いたシステムによって行われる。本手法は、3 次元位置情報を取得できる環境であれば適用可能であり、原理的には複数台のカメラ、またはデプスカメラを用いることで動作可能である。本研究では、ユーザの姿勢方向や移動に影響されない安定した認識を実現するため、実験環境として室内の天井四隅に 4 台の CCD カメラを配置した。各カメラ画像に OpenPose<sup>7)</sup>を適用し、人体の 2 次元の骨格点座標をリアルタイムで推定する。さらに、複数視点から得られた結果を統合して各関節の 3 次元座標を算出し、認識処理に用いる骨格情報を取得する。

## 3.2 アバターの姿勢の自動調整

定義開始時には、ユーザの実際の姿勢を基準として HoloLens2 に表示するアバターの初期姿勢を自動的に調整する。外部カメラから取得した骨格情報を用い、胸部を基準点として、両手首、両足首、腰、頭部の三次元座標を算出する。定義開始直後の約 5 秒間に座標を継続的に取得し、その平均値を計算することで、各骨格点の三次元平均座標を取得する。得られた平均座

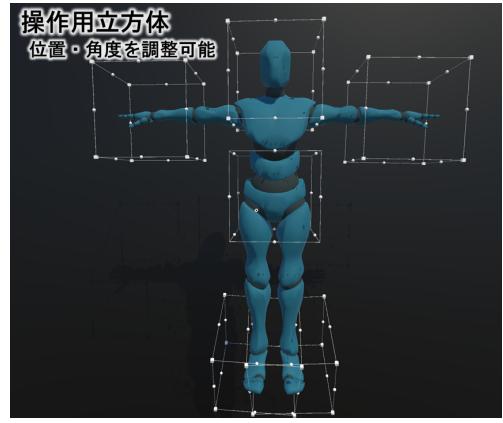


Fig. 3 Manipulation Mode

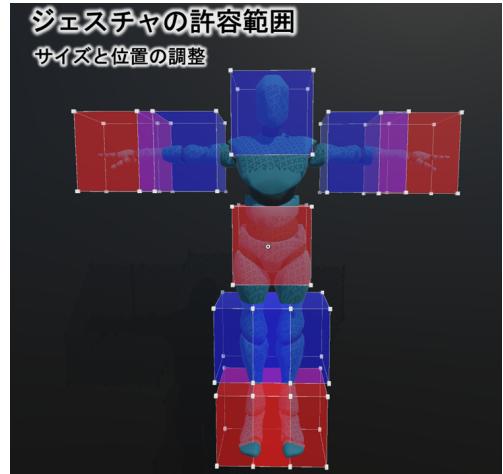


Fig. 4 Gesture Tolerance Range

標は HoloLens2 に送信され、アバターの対応部位に反映される。この処理により、ユーザは定義したいジェスチャの姿勢に近い状態から設定作業を開始でき、効率的かつ負担の少ないジェスチャ定義が可能となる。

## 3.3 ジェスチャ定義インターフェース

HoloLens2 上でのジェスチャ定義は、自動調整された等身大のアバターを直接手で操作することで行う。Fig.2(a) に示すボタンにより、操作モードを切り替えることができ、段階的にジェスチャ定義を進めることができる。例えば、左端のボタンではアバターの位置や大きさを調整可能であり、ユーザは自身の身長や作業環境に応じて操作しやすい状態に設定できる。このような初期設定により、幅広い体格や利用状況に対応した柔軟なジェスチャ定義が実現される。

### 3.3.1 ジェスチャ姿勢の設定

ジェスチャ姿勢の設定時には、Fig.3 に示すように、各主要関節の両手首、両足首、腰、頭部に操作用の立方体マニピュレータを付与している。ユーザはこれをドラッグすることで、目的の姿勢を正確に定義できる。アバターには Inverse Kinematics を適用しており、特定の部位を操作すると他の関節の動きが自動的に補完される。また、Fig.2(b) に示す部位選択ボタンを用いることで、必要な関節のみを表示・操作でき、効率的に細部調整を行うことができる。

### 3.3.2 ジェスチャの許容範囲の設定

続いて、ジェスチャ姿勢の許容範囲を設定する。Fig.4に示すように、手首、肘、膝、足首、腰、頭の部位ごとに直方体が配置されている。ジェスチャ認識時には、その範囲にすべての関節位置が収まつた場合の時に認識される。これにより、動作再現時のばらつきを許容しつつ、意図しない類似動作との混同を防ぐことができる。ユーザは直方体の位置や大きさを自由に調整可能であり、柔軟な認識条件を定義可能である。また、ここでもFig.2(c)に示す部位ごとの選択ボタンを利用することで、特定の部位に集中した調整が可能となる。

### 3.3.3 ジェスチャ座標の送信

HoloLens2で設定されたジェスチャ情報は、TCP/IP通信を用いてジェスチャ認識システムへ送信される。送信されるデータは、ジェスチャ操作の許容範囲を表す直方体の頂点の座標であり、これらはアバターの胸部を基準点とした相対座標として算出される。ジェスチャ認識モジュールでは、この情報を用いて対象関節が定義範囲内に収まっているかを照合し、ジェスチャを判定する。この仕組みにより、ユーザは定義後にHoloLens2を装着することなく操作を実行可能となり、実用に適した柔軟な操作を提供する。

## 3.4 ジェスチャ認識インターフェース

本システムにおけるジェスチャ認識は、外部カメラから取得される骨格情報を用いて行う。認識対象とする関節は、両手首・両肘・両膝・両足首・腰・頭部の計10部位であり、各関節の三次元座標を解析に用いる。さらに、利用者の一時的な動きによる誤判定を避けるため、一定時間以上動作が停止したと検出された場合にのみ判定処理を開始する設計とした。

ジェスチャ判定の基準となる座標系は、ユーザの体格差や立ち位置の違いを補正する目的で相対座標系として定義する。具体的には、胸部を原点とし、両肩を結ぶ直線をY軸、体幹の上下方向をZ軸、両者の外積でX軸を構成する。相対座標系の向きは、ジェスチャ中の姿勢変化に対して一貫した認識を維持するため、静止時から一定以上の移動が検知されるまで更新しない。ただし、原点は胸部の移動に追従するため、着座姿勢など原点の位置が変化するジェスチャであっても認識は適切に機能する。最終的な認識処理では、変換後のすべての関節座標が事前に設定された許容範囲内に収まっているかを照合し、収まっていると確認されたときにジェスチャとして認識される。

## 4. 実験

### 4.1 実験概要

提案システムの有効性を評価するため、20代の男女12名を対象として2種類の実験を実施した。評価指標は、ジェスチャ操作の所要時間、操作精度、および使用感に関する主観的評価である。

#### 4.1.1 実験1の概要

実験1では、共通の2種類のジェスチャでの比較実験を行った。本実験の目的は、ジェスチャの定義主体がジェスチャ実行時のパフォーマンスおよび主観的作業負荷に与える影響を比較検証することである。被験者は、Yポーズと着座姿勢の2種類の共通ジェスチャを、以下の2条件で実行させた。一つは、実験者が一

般的な姿勢に基づいて定義したジェスチャ、もう一つは、被験者自身が本システムを用いて定義したジェスチャである。実験者が定義したジェスチャについては、一般的な動作イメージに沿うよう設定し、全関節に対する許容範囲を一辺35cmの立方体として統一した。

評価指標として、タスク完了時間と操作精度を定量的に計測した。タスク完了後の主観的作業負荷はNASA-TLX<sup>8)</sup>を用いて評価した。NASA-TLXは、知的・知覚的欲求、身体的欲求、努力、タイムプレッシャー、フラストレーション、作業成績という6つの下位尺度で作業負荷を多角的に評価する手法である。各尺度は0から100までのアナログ尺度で評価され、それらの重みづけ平均によって総合的な作業負荷が算出される。本実験では、ジェスチャによる機器操作そのものに対する作業負荷を評価対象とした。

### 4.1.2 実験2の概要

実験2では、被験者自身が自由にジェスチャを定義し、それを用いてタスクを実行する実験を行った。本実験の目的は、提案システムにおける定義から実行までの一連のプロセスのユーザビリティを総合的に評価することである。被験者は本システムを用いて任意のジェスチャを定義し、直後にそのジェスチャを用いてジェスチャ認識を遂行した。評価指標として、タスク完了時間と操作精度を計測した。

さらに、システム全体のユーザビリティを評価するために、System Usability Scale (SUS)<sup>9)</sup>を用いた。SUSは、システムの使いやすさを定量的に評価するアンケート手法であり、以下の10項目から構成される。

- 1: このシステムを頻繁に使用したい。
- 2: このシステムは必要以上に複雑だと思う。
- 3: このシステムはシンプルで使いやすい。
- 4: このシステムを使うにはテクニカルサポートが必要。
- 5: このシステムはスムーズに機能し、連携がとれていると思う。
- 6: システムには不規則な点が多いと思う。
- 7: ほとんどの人がこのシステムをすぐに習得できると思う。
- 8: このシステムは時間がかかると思う。
- 9: このシステムを使っていると、自信が持てる。
- 10: このシステムを使い始める前に学ぶべきことはたくさんあると思う。

各質問は「強く同意する」から「強く同意しない」までの5段階リッカート尺度で回答される。スコアは規定の換算式に基づき0~100点の範囲で算出され、高いほどユーザビリティが良好であることを示す。本実験では、HoloLens2を用いたジェスチャ定義から認識に至る一連の操作全体を評価対象とした。

### 4.2 実験条件

各ジェスチャタスクにおいて、被験者は1つのジェスチャにつき3回の試行を行った。全ての実験において、ジェスチャ認識フェーズはHoloLens2を装着せずに行われた。評価指標として操作時間と操作精度を定義した。操作時間はタスク開始の合図からシステムがジェスチャを正しく認識するまでの時間として計測した。また、操作精度は3回の試行における成功率とし、タスク開始から10秒以内に認識されなかった試行は操作ミスとみなした。

**Table 1** Experiment 1 Results

	操作時間	認識失敗割合
Y ポーズ (実験者が定義)	2.71 s	30 %
Y ポーズ (ユーザが定義)	2.75 s	3 %
着席姿勢 (実験者が定義)	3.60 s	61 %
着席姿勢 (ユーザが定義)	4.32 s	42 %

**Table 2** Average NASA-TLX Score

	実験者が定義	ユーザ定義
知的・知覚的欲求	22.1	17.8
身体的欲求	14.1	9.9
タイムプレッシャー	11.3	7.5
努力	26.3	20.8
フラストレーション	14.0	8.6
作業成績	51.3	29.4
重みづけ平均値	25.9	18.0

### 4.3 実験結果・考察

#### 4.3.1 実験1の結果

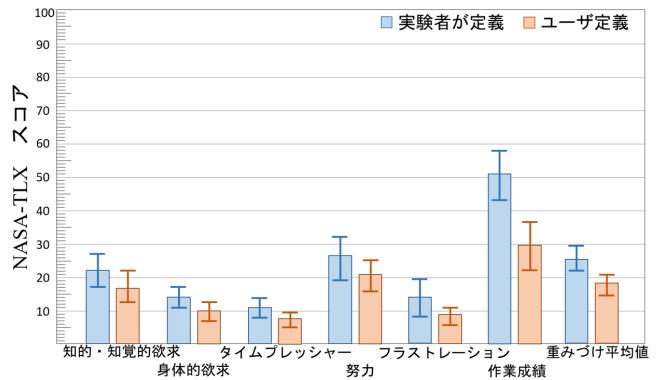
実験1における操作時間と失敗率の結果を、Table 1に示す。Y ポーズでは、実験者が定義したジェスチャの平均操作時間は 2.71s、平均失敗率は 30% であった。一方、ユーザが定義したジェスチャでは平均操作時間が 2.75s、平均失敗率は 3% であった。同様に、着座姿勢では、実験者定義の場合に平均操作時間 3.60s・失敗率 61% であったのに対し、ユーザ定義では平均操作時間 4.32s・失敗率 42% であった。これらの結果から、ユーザ定義によって操作時間はわずかに増加する傾向が見られたものの、失敗率は大幅に改善されることが確認された。

主観的作業負荷を評価するために実施した NASA-TLX の結果を、Table 2 および Fig.5 に示す。6つの下位尺度の全ての項目において、ユーザが定義したジェスチャが実験者が定義したジェスチャを下回った。また、重みづけ平均値も、実験者が定義したジェスチャの 25.9 に対し、ユーザが定義したジェスチャでは 18.0 となり、ユーザ自身が定義したジェスチャを用いることで主観的作業負荷が有意に低減することが示された。

#### 4.3.2 実験1の考察

全体を通して、ユーザ定義ジェスチャが操作成功率の向上と作業負荷の軽減に寄与することが確認された。一方で、操作時間がわずかに増加した要因としては、現在の定義インターフェースの仕様が影響した可能性がある。本システムでは認識対象の全ての関節が許容範囲内に収まることを条件としているため、被験者が定義時に一部の関節を境界付近に設定した場合、ジェスチャ再現時の微小なズレが認識遅延を引き起こし、結果的に操作時間を増加させたと考えられる。この問題に対しては、許容範囲設定を支援するインターフェースの追加や、誤差を考慮した認識アルゴリズムの導入が有効な改善策となり得る。

また、着座姿勢が Y ポーズと比較して全体的に失敗率が高い傾向が見られた。これは、Y ポーズが主に両

**Fig. 5** NASA-TLX Score**Table 3** Experiment 2 Results

	操作時間	認識失敗の割合
自由なジェスチャ	2.77 s	17 %

腕の姿勢を再現するだけで達成可能であるのに対し、着座姿勢は足や腰といった普段あまり意識されない部位を含む全身の正確な姿勢再現を必要とするため、ユーザにより高い認知的・身体的負荷を課したことが原因であると推察される。

#### 4.3.3 実験2の結果

実験2における、被験者が自由に定義したジェスチャでの操作精度および時間の結果を、Table 3 に示す。平均操作時間は 2.77s、平均失敗率は 17% であった。一方、SUS を用いたユーザビリティ評価では、平均スコアが 64.0 となり、一般的な平均値である 68.1<sup>10)</sup> を下回った。

#### 4.3.4 実験2の考察

実験2では被験者が定義したジェスチャは単純なポーズから複雑な動作まで多岐にわたった。このような多様なジェスチャに対して平均失敗率を 17% に抑えられた点は、本システムの自由度の高さと汎用性を示唆する結果といえる。一方で、SUS の平均値を下回ったことから、システムのユーザビリティには課題が残ることが明らかとなった。特に、質問項目2「このシステムは必要以上に複雑だと思う」および項目4「このシステムを使うにはテクニカルサポートが必要」で低評価が目立った。これは、ジェスチャ定義において操作対象の部位や設定パラメータが多く、タスクが煩雑化していたことや、各操作が結果に与える影響を直感的に把握しにくかったことが要因と考えられる。今後は、操作対象を必要最小限に限定する、あるいは姿勢に応じて許容範囲を自動調整する機能を導入するなど、より直感的で簡便なインターフェース設計を検討する必要がある。

### 5. 結言

本研究では、ユーザが任意に定義可能な全身ジェスチャを用いた機器操作システムを提案し、その有効性を検証した。具体的には、HoloLens2 を用いたアバター操作によりユーザ自身が自由にジェスチャを定義できるインターフェースを構築し、さらに外部カメラとの統合によって柔軟かつ高精度なジェスチャ認識を実現した。

評価実験の結果、多様なユーザ定義ジェスチャを認識可能であること、またユーザ自身が定義したジェスチャを用いることで主観的作業負荷が有意に低減することを示した。一方で、ジェスチャ定義プロセスの複雑さに起因してユーザビリティに課題が残ることも明らかとなった。

今後は、より直感的かつ簡便にジェスチャ定義を行える仕組みの導入に加え、実空間のグローバル座標系との統合によりユーザの位置や向きを考慮した認識を可能とするなど、実環境での利便性と拡張性を備えたシステムの実現を目指す。

## 参考文献

- [1] 望月勇伸, Sarthak Pathak, 梅田和昇: HoloLens2 を用いたユーザ定義のジェスチャによる機器操作システムの基礎検討, 第 43 回日本ロボット学会学術講演会, 3B1-05 (2025).
- [2] N. Schlüsener and M. Bücker: Fast Learning of Dynamic Hand Gesture Recognition with Few-Shot Learning Models, arXiv:2212.08363 (2022).
- [3] E. Rahimian, S. Zabihi, and A. Asif: FS-HGR: Few-shot Learning for Hand Gesture Recognition via ElectroMyography, arXiv:2011.06104 (2022).
- [4] A. Schäfer, G. Reis, and D. Stricker: AnyGesture: Arbitrary One-Handed Gestures for Augmented, Virtual, and Mixed Reality Applications, Applied Sciences, 12.4, p. 1888 (2022).
- [5] J. Shen et al.: Gesture Spotter: A Rapid Prototyping Tool for Key Gesture Spotting in Virtual and Augmented Reality Applications, in IEEE Transactions on Visualization and Computer Graphics, 28.11, pp. 3618–3628 (2022).
- [6] Adobe: Y Bot Mixamo, Accessed on 23.09.2025, (2025), URL: <https://www.mixamo.com/>.
- [7] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh: OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields, in IEEE Transactions on Pattern Analysis and Machine Intelligence, 43.1, pp. 172–186 (2021).
- [8] S. G. Hart and L. E. Staveland: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research, Human Mental Workload, pp. 139–183 (1988).
- [9] J. Brooke: SUS—a quick and dirty usability scale, Usability Evaluation in Industry, pp. 189–194 (1996).
- [10] J. Sauro: A Practical Guide to the System Usability Scale, Measuring Usability LLC (2011).