

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

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

近年、ジェスチャ認識を用いた機器操作システムの利用が拡大しているが、多くはあらかじめ定義された動作に限定されており、操作数や応用性に制約がある。本研究では、ユーザが任意に定義したジェスチャを用いて機器操作を行うシステム構築を目的とし、その基礎的検討を行った。HoloLens 2 上に表示されるアバターの姿勢を操作することで、任意のジェスチャを視覚的・直感的に定義可能とする。定義後は、HoloLens 2 を用いずに、外部カメラによる認識のもとでジェスチャ操作が可能となるシステムの枠組みを提案する。

1. 序論

近年、ジェスチャ認識を活用したシステムは、多岐にわたる分野で導入が進んでいる。特に、IoT 技術の普及やスマートホーム化の進展に伴い、電子機器との直感的なインタラクション手法としてジェスチャ操作が注目されている。さらに、医療や産業分野においては、機器への接触が手術の遅延や感染の要因となることを防ぐため、手の動作による非接触型のインタフェースの開発が進められている [1]。また、公共施設などにおけるディスプレイ操作に関しても、利用者が身体を使って直感的に操作可能とするインタフェースの開発が進められている [2]。このように、ジェスチャ認識に対する需要は年々高まっている。

複数の機器操作をジェスチャで実現する研究としては、Yan らは家電操作のコマンドを空間に紐づけたコマンド空間を定義し、その領域内に手をかざすジェスチャを行うことで家電を制御可能とするシステムを提案している [3]。また、我々はこれまでに、Microsoft HoloLens2[4]を用いてコマンド空間を可視化・調整を可能にすることで、ユーザごとに適した柔軟な操作環境を提供する手法を提案してきた [5]。しかし、これらの手法ではコマンド空間の数に制限があるため、実装可能な操作数に上限が生じるという課題がある。加えて、家電操作に特化していることから他分野への応用が難しく、活用できる場面が限定されてしまう点も課題である。

本研究では、ユーザが任意の身体動作をジェスチャとして定義し、そのジェスチャに基づいた機器操作を操作可能とする、汎用的かつ柔軟なインタフェースの構築を目的とする。主に家庭やオフィス環境での活用を想定し、ユーザ個人の直感や身体的特性に応じた操作環境を提供することで、操作性とアクセシビリティの向上を図る。加えて、提案手法は医療現場、製造現場、公共施設などにおける非接触操作が求められる場面への展開も視野に入れている。

本論文ではその基礎検討として、HoloLens2 を用いてユーザが自由に右腕のジェスチャを定義可能な設定・認識システムを構築する。さらに、複数のジェスチャを定義した状態で操作実験を行い、認識精度および操作時間に関する評価を通じて、本手法の有効性を検証する。

2. 関連研究

ユーザビリティの高いジェスチャ認識システムを実現するため、ユーザが独自に定義可能なジェスチャ認識手法が研究されている。Vogiatzidakis らは、家庭内の複数

機器に対する空中ジェスチャ操作を対象とし、空間的自由度の高いユーザインタラクションを実現した [6]。一方で、使用されるジェスチャは、事前に複数のユーザから意見を収集して決定された固定パターンに限定されており、任意のジェスチャの登録や動的適応には対応していない。Ye らの研究では、スマートフォン上で動作する AR ベースのプロトタイピングツールを用いて、IoT 機器とのインタラクション設計を可能とする手法が提案されている [7]。ユーザは任意の位置にジェスチャを割り当てることが可能であるが、利用できるジェスチャはあらかじめ定義された 3 種類に限られている。また、Nacenta らは、ユーザ定義ジェスチャとシステム定義ジェスチャにおける記憶性の比較を行い、ユーザ定義の方が記憶しやすく、より直感的に操作できることを報告している [8]。

他にも、ロボットとのインタラクションを目的とした動的ジェスチャ認識手法 [9] やドローン操作への応用を目的としたジェスチャ認識技術 [10] も提案されており、それぞれ特定の用途における有効性が示されている。しかし、限定された環境や使用目的を前提に設計されているため、より多様な操作ニーズや生活環境に柔軟に対応するには課題が残る。本研究は、これらの課題に対処するため、ユーザが自由に定義可能で、汎用的で柔軟なジェスチャ操作インタフェースの構築を目指す。

3. 提案手法

3.1 全体構成

本システムは、ジェスチャ定義モジュールとジェスチャ認識モジュールの 2 段階で構成される。ジェスチャの定義段階では、HoloLens2 を用いてユーザが任意のジェスチャ姿勢およびその許容範囲を 3 次元空間上で設定する。定義された情報は、ジェスチャ認識モジュールに送信され、保存される。ジェスチャの認識段階では、外部カメラから取得したユーザ動作と定義されたジェスチャとの照合を行い、一致が確認された場合には対応する機器操作が実行される。このとき、ジェスチャ定義情報は保存されたデータから読み込まれるため、認識段階では HoloLens2 を装着しない状態でもジェスチャ操作が可能である。

3.2 ジェスチャ定義インタフェース

本システムでは、ユーザが HoloLens2 を装着することで、図 1 に示すように等身大のアバターを三人称視点で確認することができる。ユーザは、このアバターを視

認しながら、自身の手で直接操作用ジェスチャの設定を行う。設定操作は、図 2 に示す各種ボタンを左から順に押下することでモードが切り替わり、直感的な操作が可能である。

まず、アバター全体のスケールと配置位置を調整する。最終的な設定完了時に、ユーザがイメージしたジェスチャ姿勢とアバターを重ねて確認できるように、アバターのスケールはユーザ身長と一致させる。

次に、図 3(a) に示すように、ユーザはアバターの右手を直接ドラッグ操作することで、対象となるジェスチャ姿勢を設定する。本システムではアバターに Inverse Kinematics 機能を適用しており、手首の位置のみを操作することで、関節の自然な動作が自動生成される。これにより、関節の物理的な可動範囲を逸脱することなく、現実的かつ再現性の高い姿勢定義を、直感的な操作で実現することが可能である。

さらに、図 3(b) に示すように、ジェスチャ姿勢の許容範囲を設定する。この設定は、動作の再現性に伴う誤差や日常的な腕の動きによる微細なズレを吸収するためのものである。また、ユーザの意図に応じて許容範囲の形状を調整することで、類似動作間の誤認識を防ぎつつ、柔軟かつ直感的な操作が可能となる。具体的には、右手首および右肘の位置に対応する 3 次元立方体をアバター上に表示し、両部位がその範囲内に収まった場合にジェスチャが認識される。ユーザはこれら立方体の位置や大きさを調整することで、認識対象とする動作の空間的許容範囲を柔軟に定義することができる。

以上の設定が完了した後、右端のボタンを押下することで、定義されたジェスチャ情報がシステムに送信・保存される。

3.3 ジェスチャ座標の送信

ユーザが HoloLens2 を用いて設定したジェスチャに関する情報は、ネットワークを介してジェスチャ認識システムへ送信される。送信には TCP/IP プロトコルを用いており、比較的軽量のデータ形式により座標情報の効率的な送受信が可能である。

送信対象となるデータは、右手首および右肘に対応する許容範囲を表す直方体の頂点の座標であり、いずれもアバター胸部を基準とした相対座標で取得される。このデータは、操作用ジェスチャが空間的にどの領域で成立すべきかを定義するものであり、後述するジェスチャ認識モジュールにおいてはこの範囲との一致をもとに判定が行われる。

このように、設定段階と認識段階を明確に分離する構成とすることで、ジェスチャ設定後は HoloLens 2 を装着せずに動作可能なシステムを実現している。

3.4 ジェスチャ認識手法

ジェスチャ認識は、ユーザが実際に機器操作を行う場面で実行され、外部に設置されたカメラによってユーザの姿勢を取得することで行われる。本研究では、ユーザの向きや位置に依存せず安定した認識を実現するため、部屋の天井四隅に配置した 4 台の CCD カメラを用いて撮影を行う。取得した画像に対しては、OpenPose[11] を適用し、人体の 2 次元骨格点をリアルタイムで推定する。得られた複数視点の情報に対してステレオ視を行うこと



図 1 システムの概要

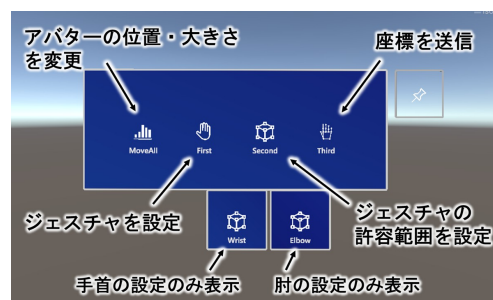


図 2 操作ボタンの概要

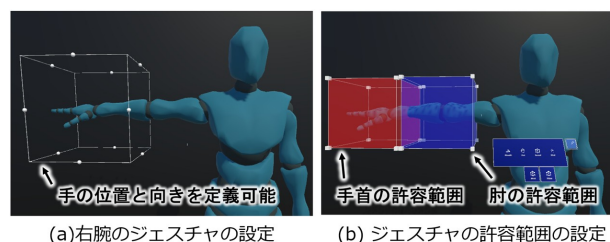


図 3 ジェスチャ設定手法

で、各骨格点の 3 次元座標を再構成する。

この処理により、胸部、両肩、右肘、右手首といった主要な部位の 3 次元座標がフレームごとに算出され、これらの情報が後続のジェスチャ判定に使用される。本システムでは、ユーザの動作が一定時間以上静止していると判定された時点でジェスチャ評価を開始する設計とし、意図しない動作による誤認識の防止を図っている。

ジェスチャ判定に際しては、まず胸部の座標を原点、両肩を結ぶ直線を Y 軸、高さ方向を Z 軸、それらの外積で X 軸を定義することで、相対座標系を構築する。この相対座標系への変換により、体格差や位置のずれといった個人差の影響を抑えた上で、肘および手首の空間的位置を安定的に評価できる。さらに、操作中の姿勢変化に対しても一貫した認識を維持するため、相対座標系の構築は動作が静止している時点でのみ実行し、ユーザが一定以上移動しない限りは肩の角度変化が生じても座標系を更新しない設計としている。最後に、変換後の肘および手首の座標が、設定時に定義された許容範囲直方体の内部に一定時間以上とどまった場合、当該ジェスチャが成立したと判定される。このようにして、本システムは、ユーザが任意に定義した動作に対して安定かつ高精度な認識を実現している。

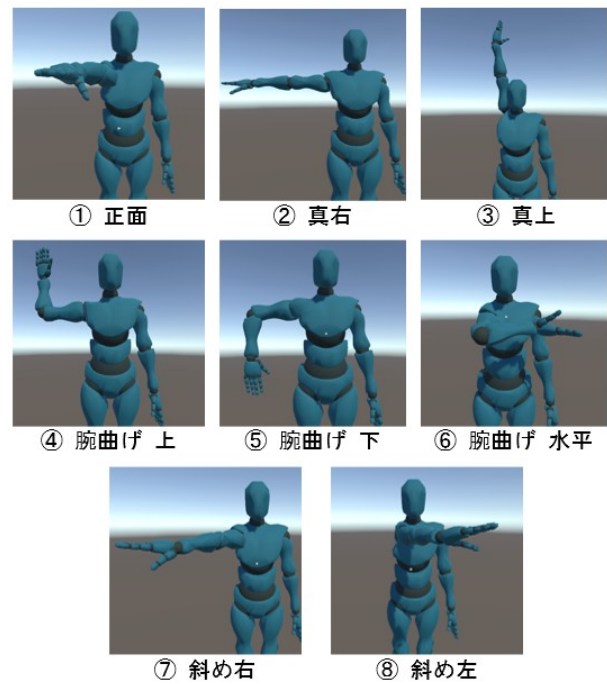


図 4 登録したジェスチャ

4. 実験

4.1 実験概要

本手法の有効性を確認するため、図 4に示す 8 種類の右腕ジェスチャを設定し、各動作に対して認識精度と操作時間の測定を行った。各ジェスチャは、動作特性に基づいて以下の 3 群に分類される。第 1 群は、腕をまっすぐに伸ばした状態で行うものであり、正面・真右・真上の 3 方向に伸ばすジェスチャである。第 2 群は、肘を曲げた状態で特定の方向に腕を向けるもので、上方向・下方向・水平方向の 3 方向に対する曲げジェスチャである。第 3 群は、特定の空間位置を指し示すような動作であり、斜め右・斜め左を指すジェスチャである。

各ジェスチャは、HoloLens2 を装着して個別に設定を行い、操作時には HoloLens 2 を装着せずに外部カメラによる認識のみで操作を行った。

4.2 実験条件

各ジェスチャについて 5 回ずつ動作を実施し、計 40 試行に対して評価を行った。認識が成功とみなされる条件は、指示から 10 秒以内に意図したジェスチャとして正確に認識されることとした。また、動作開始から認識完了までの所要時間を操作時間とする。なお、許容範囲の立方体サイズは、全ジェスチャに対して一律に 30cm 立方とした。

4.3 実験結果と考察

表 1に操作精度および時間の結果を示す。全体として、高い認識成功率と短時間での応答が得られており、提案手法の有効性が確認された。動作特性に依存せず広範なジェスチャが認識されており、特に右側の空間内で完結するジェスチャについては、非常に高い成功率が得られた。

表 1 実験結果

	平均検出時間	総失敗回数	操作成功率
①：腕を正面に伸ばす	3.77 s	0 回	100 %
②：腕を真右に伸ばす	3.05 s	0 回	100 %
③：腕を真上に伸ばす	3.64 s	0 回	100 %
④：腕曲げ 上	2.96 s	0 回	100 %
⑤：腕曲げ 下	3.20 s	0 回	100 %
⑥：腕曲げ 水平	6.16 s	3 回	40 %
⑦：斜め右	2.86 s	0 回	100 %
⑧：斜め左	3.24 s	2 回	60 %
総合	3.41 s	5 回	87.5 %

一方で、斜め左方向および水平方向への肘曲げ動作においては、他のジェスチャと比較してやや認識率が低下する傾向が見られた。これらの動作は、いずれも身体の左側に向かって腕を動かす必要があるため、姿勢に無理が生じやすく、ユーザが同一姿勢を再現する際のばらつきが大きくなったことが認識率低下の要因と考えられる。

5. 結言

本研究では、ユーザが自由にジェスチャを定義し非接触で機器を操作可能とする新たなインタフェースの基礎構築を行った。HoloLens2 を用いて姿勢およびその許容範囲を視覚的かつ直感的に定義し、定義後は HoloLens2 を装着することなく、ジェスチャのみで機器操作を実現できることを示した。

実験により、高い認識精度と良好な操作性が得られ、提案手法の有効性が確認された。一方で、空間的指標が不明瞭な動作においては、ユーザの再現性にばらつきが生じやすく、安定性に課題があることも明らかとなった。今後は、Few-Shot 学習を活用した画像ベースの補助認識手法や、ユーザ特性に応じた動的閾値調整アルゴリズムの導入を検討し、さらなる認識の安定化を図る予定である。

また、現段階では右腕の動作に限定された操作に留まっているが、今後は全身を用いたジェスチャ認識を可能にする手法を検討する。さらに、現状の身体相対座標系のみに基づくシステムを、共通空間上のグローバル座標系へと拡張し、ユーザの位置や向きを考慮した、より柔軟かつ高精度なジェスチャ定義が可能なシステムの構築を目指す。

参 考 文 献

[1] Mithun George Jacob, Juan Pablo Wachs, Rebecca A Packer, “Hand-gesture-based sterile interface for the operating room using contextual cues for the navigation of radiological images,” Journal of the American Medical Informatics Association, 2013

[2] Qiujiiao Wang, Zhijie Xie, “ARIAS: An AR-based interactive advertising system”, PLoS ONE, vol.18, no.1, e0280000, 2023

[3] Shixun Yan, Yonghoon Ji, Kazunori Umeda, “A System for Operating Home Appliances with Hand Positioning in a User-definable Command Space,” 2020

- IEEE/SICE International Symposium on System Integration (SII), pp. 366-370, 2020.
- [4] 日本マイクロソフト, “Microsoft HoloLens - Learn Microsoft,” <https://www.microsoft.com/ja-jp/hololens> (Accessed on 7/2/2025)
- [5] Yushin Mochizuki, Masae Yokota, Sarthak Pathak and Kazunori Umeda, “Visualisable and Adjustable Command Spaces for Gesture-based Home Appliance Operation System via HoloLens2,” 2025 IEEE/SICE International Symposium on System Integration (SII), pp. 1405-1410, 2025
- [6] Panagiotis Vogiatzidakis, Panayiotis Koutsabasis “Mid-Air Gesture Control of Multiple Home Devices in Spatial Augmented Reality Prototype“ Multimodal Technologies and Interaction (MTI), vol. 4, no. 3, pp. 61 , 2020
- [7] Hui Ye, Hongbo Fu, “ProGesAR: Mobile AR Prototyping for Proxemic and Gestural Interactions with Real-world IoT Enhanced Spaces“, CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, No. 130, pp. 1-14, 2022
- [8] Miguel A. Nacenta, Yemliha Kamber, Yizhou Qiang, Per Ola Kristensson, “Memorability of Pre-designed and User-defined Gesture Sets“ Proceedings of the 31st Annual ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2013), pp. 1099-1108, 2013
- [9] Qing Gao, et al., “Dynamic Hand Gesture Recognition Based on 3D Hand Pose Estimation for Human-Robot Interaction,” IEEE Sensors Journal, vol. 21, no. 15, pp. 16476-16486, 2021
- [10] Mohd Iskandar, Kishore Bingi, Rosdiazli Ibrahim, Madiyah Omar and P.Arun Mozhi Devan, “Hybrid Face and Eye Gesture Tracking Algorithm for Tello EDU RoboMaster TT Quadrotor Drone,” 2023 Innovations in Power and Advanced Computing Technologies (i-PACT), pp. 1-6, 2023
- [11] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei and Yaser Sheikh, “OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields,” in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 1, pp. 172-186, 2021