

遠隔ロボットアームシステムにおける VR映像に基づく把持作業支援

坪内 陸斗^{1,a)} 筒井 秀斗¹ 出口 秀輝¹ 渡辺 圭貴¹ 志村 魁哉¹ 林田 望海¹ Felix Jimenez²
浦野 健太¹ 米澤 拓郎¹ 河口 信夫^{1,2}

概要：遠隔ロボット操作は、災害現場や汚染環境などの危険環境、および人手不足地域における作業支援手段として重要性が高まっている。こうした実環境での作業には対象物の選択や作業手順の判断、予測困難な例外処理が求められるため、人間の認知・判断を組み込んだ遠隔操作が不可欠である。遠隔操作では、操作者が作業環境を直感的に認識できる映像フィードバックが重要であり、VR (Virtual Reality) 映像を用いた没入型インタフェースが注目されている。しかし、VR 映像では奥行き知覚に制約があり、把持操作における精密な位置合わせに操作負荷が生じやすい。この負荷を軽減するため、操作者が把持対象の選択に集中できるよう把持動作を半自動化する支援構造が求められる。本研究では、VR 遠隔ロボットアーム操作における把持タスクを対象に、操作者が必要に応じて手動操作と自動把持を明示的に切り替える「明示的切替型」の半自動支援構造を提案する。従来の共有制御や適応的共有自律がシステム主導で支援量を調整するのに対し、提案手法では自動モードへの移行・復帰を操作者が制御するため、操作者がロボットの動作根拠を常に把握でき、意図しない自律動作を防止できる。手先カメラ映像に対する物体セグメンテーションにより対象物の位置を推定し、アームの位置操作から対象物の選択へ移行するタイミングが異なる3つの対象物選択手法（手先位置選択・最近傍選択・スティック選択）を実装した。平面的な対象物を扱った予備実験では、位置合わせを簡略化した最近傍選択が完遂時間・主観的評価ともに最良の成績を示し、手先位置選択の上位互換として機能する仮説が立てられた。しかし、三次元的に物体が近接・混在する環境を用いた本実験（12名）においては結果が逆転し、手先位置選択が完遂時間および操作負荷の観点で最も安定した値を示す傾向が確認された。一方で、3手法間に統計的有意差は認められず、配置密度に応じた手法の使い分けの可能性も示唆された。

1. はじめに

遠隔ロボット操作は、人が直接立ち入れない環境や、労働力が不足する現場における作業手段として期待が高まっている。例えば、原子力施設の保守・廃炉作業では高線量環境下での弁操作や配管切断、化学プラントでは漏洩事故の際に危険環境での対応が要求され [1]、建設分野や農業分野では人手不足解消を目的とした応用が広がっている。これらの実環境における作業は、産業用ロボットが得意とする反復的な工場内タスクとは本質的に異なり、操作対象の形状・配置が毎回変化するため、作業者は対象物の選択、把持位置の決定、障害物の回避など、状況に応じた逐次的な判断を行う必要がある。こうした判断はルールベースの

自動化では網羅が難しく、人間の判断を介した遠隔操作システムが不可欠である。

遠隔操作では、操作者が作業環境を直感的に認識できる映像フィードバックが重要であり、近年はVR技術を用いた没入型の操作インタフェースが注目されている [2]。VR映像はHMD (Head-Mounted Display) を通じて操作者に立体視を提供し、従来の2次元モニタ方式では困難であった奥行き方向の距離感や対象物の立体的な形状把握を支援する [3]。一方で、VR映像による奥行き知覚は実環境での両眼立体視と比較して精度が低く [4]、エンドエフェクタの精密な位置合わせが要求される把持操作においては、操作者に対する認知的・身体的な負担が増大する。とりわけ対象物直近でのアプローチ制御は微細な操作を要するため、操作の効率と正確性の低下が課題となる。

この操作負荷を軽減するアプローチとして、人とロボットの制御入力を動的に統合する「共有制御 (Shared Control)」 [5] や、操作者の熟練度やタスク状況に応じて支援量を調整

¹ 名古屋大学大学院工学研究科
Graduate School of Engineering, Nagoya University

² 名古屋大学未来社会創造機構
Institutes of Innovation for Future Society, Nagoya University

^{a)} rikuto@ucl.nuee.nagoya-u.ac.jp

する「適応的共有自律 (Adaptive Shared Autonomy)」[6] が研究されてきた。しかし、これらの枠組みでは制御入力の統合がシステム内部で行われるため、操作者にとってロボットの動作根拠が不透明になりやすい。また、意図推定の誤りに起因して操作者の意図とは異なる動作が実行されるリスクがあり、支援の発動タイミングを操作者自身が選択する余地がないため操作者の主体性も制約される。このように、透明性・安全性・主体性の観点で課題があり、特に安全性が重視される遠隔作業環境においては実用上の大きな障壁となる。さらに、自動把持における対象物の選択方法について、位置操作から選択操作への切り替えタイミングを体系的に比較した研究は十分に行われていない。

そこで本研究では、VR 遠隔ロボットアーム操作における把持タスクを対象に、操作者が必要に応じて手動操作と自動把持を明示的に切り替える明示的切替型の半自動支援構造を提案し、実環境で動作する遠隔操作ロボットに実装する。提案手法では自動モードへの移行・復帰を操作者が制御するため、自動モード中であっても操作者が制御の主体であり続け、意図しない自律動作を防止できる。手先カメラ映像に対する物体セグメンテーションにより対象物の位置を推定し、アームの位置操作から対象物の選択へ移行するタイミングが異なる 3 つの対象物選択手法を実装した。自動把持を開始する際、操作者はアームを対象物の近くまで位置操作した後、選択操作に移行して把持対象を決定する。この位置操作と選択操作の境界は連続的に変化しうるのであり、精密な位置操作を経て選択する方式から、大まかな位置操作の後にスティック操作で選択する方式まで、操作負担の配分は段階的に異なる。本研究では、この連続的な設計空間の中から代表的な 3 点として、対象物の直上まで精密に位置合わせを行う手先位置選択、およその接近で最近傍の物体を自動選択する最近傍選択、大まかな接近後にスティック操作で候補を指定するスティック選択の 3 手法を実装し、作業効率と操作の容易さの観点から最適な設計点を探索する。

平面的な配置を用いた予備実験では、位置合わせの精度要求を緩和した最近傍選択が最も良好な結果を示した。一方、三次元的な配置を用いた本実験 (12 名) では結果が逆転し、手先位置選択が完遂時間・操作負荷の観点で安定した値を示した。3 手法間に統計的有意差は認められず、また高密度環境ではスティック選択の完遂時間の増加率が低く、配置密度による手法の適性の違いが示された。明示的切替型構造に対しては「自動化の発動を自分で制御できる安心感がある」との肯定的なフィードバックが得られた。

本研究の貢献は以下のようにまとめられる。

- 操作者が支援の発動・停止を明示的に制御する明示的切替型の半自動把持支援構造を設計し、実環境で動作する遠隔操作ロボットに実装した点
- 位置操作から選択操作への移行タイミングが異なる 3

種類の対象物選択手法を、連続的な設計空間の代表点として実装し、作業効率・操作性に基づく体系的な比較評価を行った点

- 予備実験と本実験で結果の逆転が生じたことを報告し、タスク環境の複雑さと実験設計が手法評価に与える影響を示した点

2. 関連研究

2.1 VR 映像を用いた遠隔ロボット操作

遠隔ロボット操作 (テレオペレーション) は、1940 年代後半に Goertz が放射性物質の取り扱いを目的としたマスタースレーブ型マニピュレータを開発して以降 [7]、操作者とロボットが物理的に離れた位置でも動作可能な方式が確立され [8]、宇宙探査・深海探査・医療・災害対応など幅広い分野で実用化が進んでいる [1]。

遠隔操作では、操作者が作業環境を正確に認識するための映像フィードバックが重要であり、近年は VR 技術を用いた没入型インタフェースの研究が増加している [3]。2 次元モニタを用いた従来方式では、奥行き方向の距離感の把握が困難であり、精密な位置合わせ操作において操作時間の増加やエラー率の上昇が報告されている [11]。これに対し、HMD を用いた VR 映像は操作者に立体視を提供し、対象物の三次元的な位置関係の把握を容易にする。また、ロボットに搭載したステレオカメラからの映像を HMD に投影する方式に加え、RGB-D センサから三次元点群を復元して VR 空間に描画する方式 [10] も提案されており、いずれも操作者の空間認識を向上させる手段として研究されている。

Whitney ら [9] は、消費者向け VR ハードウェアと ROS (Robot Operating System) を統合する遠隔操作フレームワーク “ROS Reality” を開発し、従来のモニタベース方式と比較して SUS (System Usability Scale) スコアの有意な向上を報告した。Lipton ら [2] は、VR 映像に基づく遠隔ロボット操作フレームワークを提案し、操作者の没入感と空間認識の向上を示した。

しかしながら、VR 映像の奥行き知覚は、輻輳・調節矛盾などの視覚的制約により、実環境での両眼立体視と比較して精度が低下する問題が知られている [4]。このため、特にエンドエフェクタが対象物に接近する最終段階において、微細な位置合わせに困難が生じる。この課題に対し、Arevalo Arboleda ら [11] は、奥行き知覚の制約を補うために拡張現実 (AR) による視覚の手がかりを付加する手法を提案し、把持精度の改善を確認している。加えて、長時間の HMD 装着に伴う眼精疲労や VR 酔いなどの生理的負荷も課題として指摘されている [3]。本研究では、これらの VR 固有の課題のうち、把持時の精密な位置合わせに要する操作負荷の軽減に焦点を当てる。

表 1 関連研究アプローチと本研究の比較

比較項目	共有制御	適応的共有自律	明示的切替型（本研究）
支援の発動主体	システム	システム	操作者
仲裁方式	policy blending	動的係数調整	モード切替
意図推定	必要	必要	不要
透明性	低～中	低～中	高
操作者の主体性	制約あり	制約あり	制約なし
対象物選択の設計比較	なし	なし	3 手法を比較

2.2 共有制御と適応的共有自律

人とロボットの制御を統合する枠組みとして、共有制御の研究が広く行われている [5], [12]. 共有制御では、操作者の入力とシステムの自律動作を一定の仲裁方策に基づいて融合し、操作者が制御の主体でありながらシステムの支援を受けるという構造を実現する. Dragan ら [5] は、操作者の行動観測から目的関数を推定する逆強化学習を用いてユーザの意図を逐次推定し、ロボットの動作を操作者の意図に合致させるよう補助する policy blending の枠組みを提案した. Javdani ら [13] は、ユーザの目標に関する不確実性を部分観測マルコフ決定過程 (POMDP) として定式化し、hindsight optimization により目標分布全体に対する最適支援行動を近似する手法を示した. Zhu ら [12] は、障害を持つユーザの遠隔操作を対象とし、共有制御による操作負荷の軽減効果を報告している.

共有制御をさらに発展させた枠組みとして、適応的共有自律がある [6]. この手法では、操作者のパフォーマンスやタスクの進行状況に応じて人間とロボットの制御配分を動的に変化させる. 例えば、操作者の習熟度が低い段階ではシステムの支援比率を高め、熟練するにつれて操作者の制御比率を高めるといった適応が可能となる. Reddy ら [14] は、深層強化学習を用いて操作者の目標やモデルが未知の場合でも支援を行える枠組みを提案し、適応的な支援の実現可能性を示した.

これらの枠組みは操作支援の効果が報告されている一方で、いくつかの課題が存在する. 第一に、制御入力の統合がシステム内部で行われるため、操作者にとってロボットの動作根拠が不透明になりやすい. 第二に、意図推定の誤りに起因して、操作者の意図とは異なる動作が実行されるリスクがある. 第三に、支援の発動タイミングを操作者が選択できないため、操作者の主体性が制約される. 特に安全性が重視される遠隔作業環境においては、これらの課題は実用上の大きな障壁となりうる.

2.3 把持動作の自動化に関する研究

ロボットによる把持動作の自動化は、遠隔操作とは独立した研究分野としても広く研究されてきた. 深層学習に基づく把持点検出では、RGB 画像や深度画像から把持可能な位置と姿勢を推定する手法が多数提案されている [15]. Newbury ら [16] は深層学習による把持合成手法を体系的

にレビューし、サンプリングベース、直接回帰、強化学習、exemplar ベースの 4 つの方法論に分類している. また、物体のセグメンテーションと姿勢推定を組み合わせた把持計画や、点群データから直接把持方策を学習するアプローチなども報告されている [15].

遠隔操作の文脈においても、自動把持機能を部分的に導入して操作者の負荷を軽減する試みが存在する. Lipton ら [2] のフレームワークでは VR 環境内で操作者が対象物に接近した後のファインアプローチをシステムが補助する構成が示されている. また、対象物を操作者が指定した後にアプローチ軌道を自動生成する支援システムも報告されている [1]. これらの研究は本研究と共通する問題意識を持つが、位置操作から選択操作への移行タイミングの設計を体系的に比較した研究は十分に行われていない.

2.4 本研究の位置づけ

表 1 に、関連する研究アプローチと本研究の比較を示す.

共有制御 [5], [13] や適応的共有自律 [6], [14] がシステム主導で支援量を調整するのに対し、本研究では操作者が支援の発動タイミングを明示的に決定する明示的切替型の構造を採用する. この方式では、操作者は手動モードでロボットを操作し、把持が必要なタイミングで自らの判断により自動モードに切り替える. 自動モードにおいても、操作者は動作の開始・中止を制御する権限を持つため、意図しないロボット動作を防止できるとともに、システムの振る舞いを操作者が常に把握できる.

さらに本研究では、自動把持の対象物をどのように決定するかという対象物選択の設計に着目する. 自動把持を開始する際、操作者はアームの位置操作によって対象物に接近した後、選択操作に移行して把持対象を決定する. 対象物の直上まで精密にアームを運ぶ方式から、大まかに接近した後スティック操作で対象を指定する方式まで、位置操作と選択操作の間の負担配分は連続的に変化しうる. 本研究では、この連続的な設計空間の中から代表的な 3 点として 3 種類の選択手法を実装・比較し、作業効率と操作の容易さの観点から適切な設計点を明らかにする. この観点は、既存の共有制御研究において十分に検討されておらず、本研究の主たる新規性である.

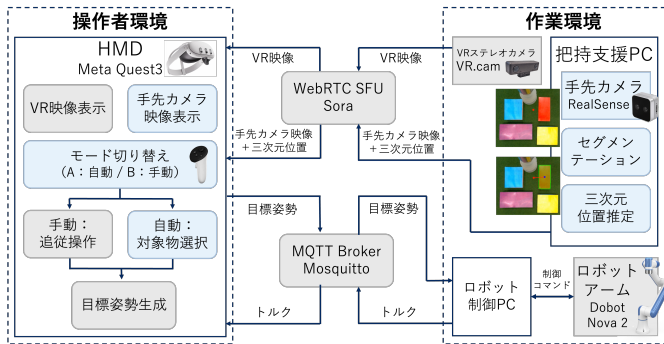


図 1 システム全体構成

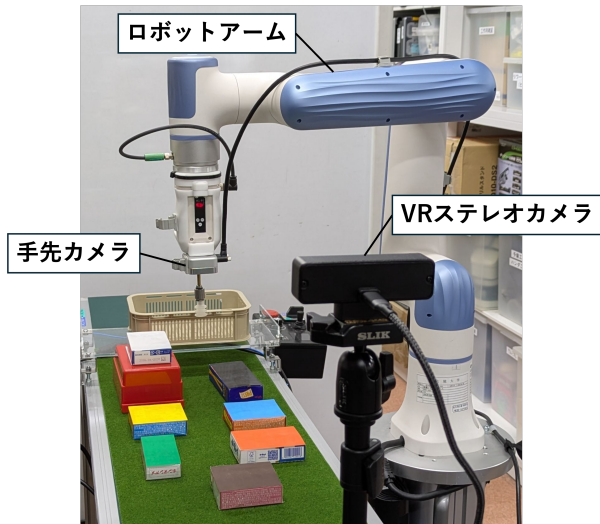


図 2 作業環境の外観

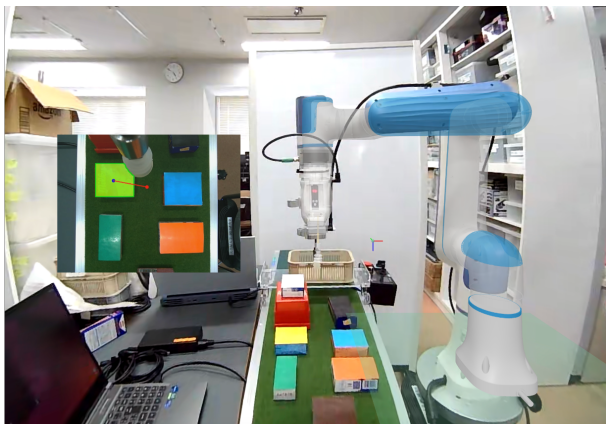


図 3 HMD を通じた操作者視点映像

3. 提案システム

3.1 全体構成

本研究が拡張対象とする VR 遠隔ロボットアーム操作システムの全体構成を図 1 に示す。また、作業環境を図 2 に、HMD を通じて提示される映像を図 3 に示す。本システムは、HMD・ロボットアーム・手先カメラ・把持支援モジュールの 4 要素で構成される。

本システムの基盤となる VR 遠隔操作フレームワークは

Tsutsui ら [17] が構築した既存システムを使用する。既存システムは HMD による VR 映像表示、手動操作モードによるロボットアームの追従制御、および WebRTC を介した映像伝送機能を備える。本研究ではこの基盤に対し、手先カメラ映像を用いた物体セグメンテーション、対象物の三次元位置推定、3 種類の対象物選択手法、および自動把持シーケンスを新たに実装・統合した (図 1 の青色部分)。

操作者は HMD を装着し、ロボット手先に搭載されたカメラからの映像をリアルタイムで観察しながら、VR コントローラでロボットアームを操作する。本システムの通信は 2 系統で構成される。手先カメラと HMD 間の通信には WebRTC (Web Real-Time Communication, 実装には Sora SDK を使用) を使用し、カメラ映像の伝送に加え、データチャンネルを介して対象物の三次元位置推定結果を HMD 側に送信する。HMD とロボット間の通信には MQTT プロトコル (Message Queuing Telemetry Transport) を使用し、エンドエフェクタの目標姿勢をロボットに送信する。手動モードでは、VR コントローラの移動増分に対してモーションフィルタを適用した上で目標変位を生成し、自動モードでは、受信した三次元位置推定結果に基づき HMD 内で自動把持の動作軌道を生成した上で、目標姿勢を MQTT 経由でロボットに送信する。把持支援モジュールは、手先カメラ映像を入力として物体セグメンテーションおよび対象物の位置推定を行い、セグメンテーション結果の重畳映像を Sora 経由で HMD に描画する。

VR 側アプリケーションは Next.js フレームワーク上で A-Frame (Three.js ベース) および WebXR を用いて実装した。ロボット制御には Dobot 公式 API (dobot.api) を使用し、把持支援モジュールは Python で実装した。

3.2 ハードウェア構成

表 2 に本システムのハードウェア構成を示す。ロボットアームには Dobot Nova 2 を使用した。Nova 2 は 6 自由度の協働ロボットであり、最大リーチ 625 mm, 可搬重量 2 kg の仕様を持つ。なお、本システムはロボットアームの制御インタフェースを抽象化しており、Nova 2 以外のロボットアームにも対応可能な設計となっている。本研究では Nova 2 を用いて開発・評価を行った。エンドエフェクタには真空吸引グリップを採用した。吸着式グリップは、対象物の上面に接触するだけで把持が可能であり、把持時のアプローチ方向が鉛直下向きに限定されるため、自動把持における制御の簡素化に寄与する。

手先カメラには Intel RealSense D405 (RGB-D, RGB-Depth) を搭載した。RealSense はエンドエフェクタの直近に固定されており、カメラ光軸はエンドエフェクタの鉛直下方向とほぼ一致する。RGB 画像に加えて深度情報を取得可能であり、把持支援モジュールにおける対象物の三

表 2 ハードウェア構成

要素	仕様
ロボットアーム	Dobot Nova 2 (6 軸協働ロボット) 最大リーチ: 625 mm 可搬重量: 2 kg 繰返し精度: ± 0.05 mm
エンドエフェクタ	真空吸引グリップ (吸着式)
手先カメラ	Intel RealSense D405 解像度: 640×480, 30 fps
VR ステレオカメラ	VR.cam 02 解像度: 3840×1920, 30 fps 視野角: 160°
HMD	Meta Quest 3

表 3 VR コントローラのボタン割り当て

操作	機能
右トリガー (押下中)	手動追従動作
右グリップ (押下中)	エンドエフェクタの吸着 ON
A ボタン	自動把持シーケンス開始
B ボタン	手動モードへ復帰
右サムスティック	対象物候補の切替 (スティック選択時)

次元位置推定に使用する。画像取得は 30 fps で行い、把持支援モジュールの処理系は 15 fps で動作する。

HMD には Meta Quest 3 を使用した。Meta Quest 3 はスタンドアロン型の HMD であり、VR コントローラとのワイヤレス接続により操作者の身体的拘束を最小限に抑える。コントローラのボタン割り当てを表 3 に示す。

3.3 操作モードと安全設計

本システムは手動モードと自動モードの 2 つの操作モードを持つ。手動モードでは、VR コントローラの位置変化にロボットアームを追従させ、操作者が自由にアームを操作する。手動モードではデッドマンスイッチ方式を採用しており、右トリガーを押下している間のみロボットが追従動作を行う。トリガーを離した瞬間にロボットは目標値の更新を停止し、動作を中断する。

後述する対象選択手法で決定した対象物に対し、A ボタンを押すと把持シーケンスが自動的に進行する。B ボタンを押せばシーケンスの任意の段階で即座に手動モードへ復帰でき、手動操作時の VR コントローラ入力からロボット動作までの遅延は約 200 ms である。

3.4 対象選択手法

本研究では、アームの位置操作から対象物の選択への移行タイミングが異なる 3 種類の対象物選択手法を設計した (図 4)。各手法はいずれも、第 3.5.1 項で述べる物体セグメンテーションにより生成されたマスク領域を用いて対象物を選択する。自動把持を開始する際、操作者はまずアーム



図 4 対象選択手法の概念図 (左: 手先位置, 中: 最近傍, 右: スティック)

の位置操作によって対象物に接近し、その後選択操作に移行して把持対象を決定する。この位置操作と選択操作の境界は連続的に変化するものであり、3 手法はこの連続的な設計空間における代表的な 3 点として位置づけられる。以下に各手法を説明する。

3.4.1 手先位置選択

エンドエフェクタの鉛直下方投影点が物体のセグメンテーションマスク領域内に含まれる場合にのみ、その物体を把持対象として選択する手法である。操作者はエンドエフェクタを対象物の直上まで正確に移動させる必要があり、位置操作の精度要求が 3 手法中最も高い。一方で、選択操作は不要であり、マスク内にエンドエフェクタが位置する状態で A ボタンを押すだけで自動把持が開始される。対象物が存在しない位置では自動把持が開始されないため、誤選択のリスクは低い。VR 環境での精密な位置操作が要求される分、操作者の認知的・身体的負担が増大する。

3.4.2 最近傍選択

エンドエフェクタの現在位置から画像座標上で最も近い物体を自動的に把持対象として選択する手法である。距離はマスク重心とエンドエフェクタ投影点の間のユークリッド距離で計算する。なお、エンドエフェクタの投影点がマスク内に含まれる物体が存在する場合は、その物体を優先的に距離 0 として扱う。操作者はエンドエフェクタをある程度対象物に近づけた状態で A ボタンを押すだけでよく、手先位置選択と比較して位置操作の精度要求が緩和される。一方で、複数の物体が近接する環境では意図しない物体が選択される誤選択のリスクが生じる。

3.4.3 スティック選択

手先カメラ映像上でセグメンテーションにより検出された対象物の候補をエンドエフェクタからの距離が近い順に番号を付与して操作者に提示し、VR コントローラの右サムスティックを左右に倒して n 番目に近い候補を選択する手法である。フレーム間での候補順位の安定化には ByteTrack[18] によるトラッキングを使用し、各候補に対して番号と輪郭線を重畳表示する。操作者はスティックで所望の候補を選択した状態で A ボタンを押す、自動把持を開始する。

3 手法中、位置操作の精度要求が最も低く、エンドエフェ

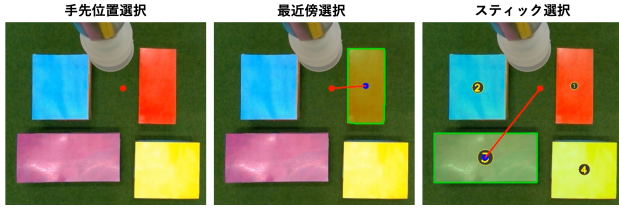


図 5 FastSAM による物体セグメンテーション例

クタを大まかに対象物群の近くまで移動させれば選択操作に移行できる。一方で、候補の視覚確認とスティック操作による認知的・操作的負荷が増加し、特に候補数が多い環境では選択に時間を要する。

3.5 自動把持アルゴリズム

3.5.1 物体セグメンテーション

手先カメラから取得した RGB 映像に対し、FastSAM[19]を適用して物体領域のマスクを生成する。FastSAM は Segment Anything Model (SAM) [20] を高速化した軽量モデルであり、リアルタイム処理に適している。本システムでは FastSAM-s モデルを使用し、入力画像を 256×256 にリサイズして推論を行う。推論パラメータは信頼度閾値 $\text{conf} = 0.3$ 、IoU 閾値 (Intersection over Union) $\text{iou} = 0.7$ に設定した。セグメンテーション結果に対しては以下のフィルタリングを適用し、把持対象として不適切な領域を除外する。

- 背景領域の除外：画像全体の面積の 30% を超える大面積マスクを除外
- ロボット自身の映り込み領域を含むマスクを除外

セグメンテーション例を図 5 に示す。

3.5.2 対象物の三次元位置推定

対象物の三次元位置推定は、RealSense の RGB-D 情報を用いて行う。セグメンテーションにより検出されたマスク領域の重心座標 (u, v) を求め、同座標における深度値 d を RealSense の深度フレームから取得する。画像座標 (u, v) と深度 d から、RealSense SDK の `rs2_deproject_pixel_to_point` 関数により、カメラ内部パラメータ（焦点距離、主点）を用いてカメラ座標系における三次元座標 (x_c, y_c, z_c) を算出する（式 (1)）。

$$\begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} = d \cdot \begin{pmatrix} (u - c_x)/f_x \\ (v - c_y)/f_y \\ 1 \end{pmatrix} \quad (1)$$

ここで f_x, f_y は焦点距離、 c_x, c_y は主点座標であり、深度フレームのストリームプロファイルからリアルタイムに取得する。

手先カメラはエンドエフェクタの直近に固定されているため、エンドエフェクタ中心のカメラ座標系における三次元座標と対象物の三次元座標の差分 $(\Delta x, \Delta y, \Delta z)$ を把持

目標位置への移動量として算出する。この移動量は Sora のデータチャンネルを介して HMD 側に送信され、HMD 内で自動把持の動作軌道を生成する際の位置指令に使用される。

3.5.3 自動把持シーケンス

自動把持の開始後、ロボットは以下のシーケンスで把持動作を実行する。シーケンスの任意の段階で操作者が B ボタンを押した場合は直ちに動作を中断し手動モードへ復帰する。

- (1) **対象物の決定**：選択手法に基づき把持対象を決定する。
- (2) **XY 位置合わせ**： $\Delta x, \Delta y$ に基づき、エンドエフェクタを対象物の鉛直上方に移動する。
- (3) **高速降下**：対象物の推定位置から 10 cm 上方の位置まで高速に降下する。
- (4) **低速降下 + 接触検知**：1 mm/step の超低速で降下しながら、各関節のトルク値を監視する。接触判定にはハイブリッド方式を採用し、定常状態からのトルク総変化量の L2 ノルムが閾値 $\tau_{\text{total}} = 2.0 \text{ N}\cdot\text{m}$ を超え、かつ前ステップからの急変量の L2 ノルムが閾値 $\tau_{\text{spike}} = 1.1 \text{ N}\cdot\text{m}$ を超えた時点で接触と判定する。深度情報に基づく降下限界として推定深度の 120% の距離に達した場合も降下を停止する。
- (5) **上昇**：対象物を把持したまま所定の高さまで上昇する。
- (6) **搬送・格納**：把持した対象物を指定のかご上方まで搬送し、吸着を解除してかごへ格納する。
- (7) **初期姿勢復帰**：ロボットアームを初期姿勢に戻し、手動操作へ復帰する。

ステップ 3 と 4 の分離により、対象物への接近は高速に行いつつ、接触直前の微細な制御では低速化とトルク監視を組み合わせ、安全な把持を実現する。

4. 評価実験

4.1 予備実験による基礎検討

本実験環境を用いた評価に先立ち、提案する自動把持支援の有効性を確認するための予備実験を実施した。この段階のシステムでは、深度センサを用いた高さ推定機能が未実装であり、把持時の降下高さを固定としていた。また、通信処理等に起因する映像・操作の遅延が存在する状態であった。

予備実験では被験者 12 名を対象とした被験者内計画を採用し、机上に平置きされたトランプのカードを指定順に把持するタスクにて 3 手法を比較した。その結果、タスク完遂時間および主観的な使いやすさ評価の双方において、最近傍選択、手先位置選択、スティック選択の順で良好な成績を示す傾向が得られた（統計的有意差はなし）。この結果から、位置合わせの精度要求を緩和した最近傍選択が、手先位置選択の操作負荷を軽減する上位互換として機能するという仮説を立てた。

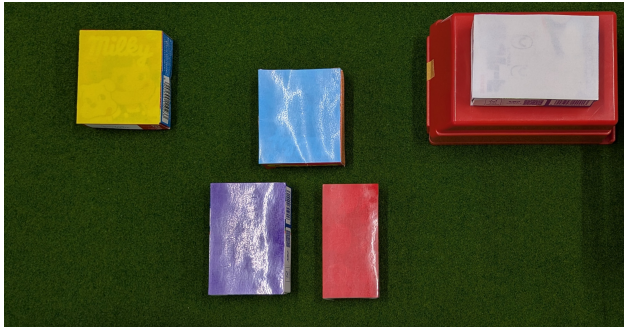


図 6 本番 1 の配置環境

しかし、この予備実験は対象物が二次元的に配置され、かつ把持高さが一定という平面的なタスクであった。より実際の作業環境に近い条件における手法の特性を検証するため、三次元的な配置と高さの変化（台座の有無）を含む本実験環境を構築し、次節以降で詳述する評価実験を実施した。

4.2 実験目的

前節の予備実験で得られた仮説を検証するため、本実験では、実装した 3 手法の作業効率・操作性・操作負荷の比較を目的とする。一般的な遠隔把持操作を想定した条件のもとで実験を設計し、配置密度の異なる 2 つの難易度環境を用いて手法間の傾向を検証する。

4.3 被験者

12 名（男性 11 名、女性 1 名、平均年齢 21.2 歳、標準偏差 1.2 歳）が被験者として実験に参加した。全被験者は VR 遠隔ロボットアーム操作の経験を持たず、正常な視力を有していた。

4.4 実験設計

本実験は被験者間計画を採用し、12 名の被験者を 3 手法に各 4 名ずつランダムに割り当てた。被験者間計画の採用により、手法間の学習効果の影響を排除した。

実験は操作習熟を目的とした練習セッションと、評価対象となる本番セッションの計 4 セッションで構成し（表 4）、各セッション間には休憩を設けた。

タスクは、テーブル上に配置された色の異なる箱を指定された順序でロボットアームにより把持する作業とした。箱の配置は全被験者で統一し、セッションごとに異なる配置を使用した。把持対象および把持順序は実験者が事前に指定し、被験者は VR 映像を通じて指示を確認しながらタスクを遂行した。

本番 1・本番 2 の配置環境を図 6 および図 7 に示す。両セッションは物体の配置密度と台座による高さ変化の程度が異なり、本番 1 は中密度かつ台座 1 個に箱 1 個を配置した中難易度環境、本番 2 は高密度かつ台座 3 個に計 4 個の

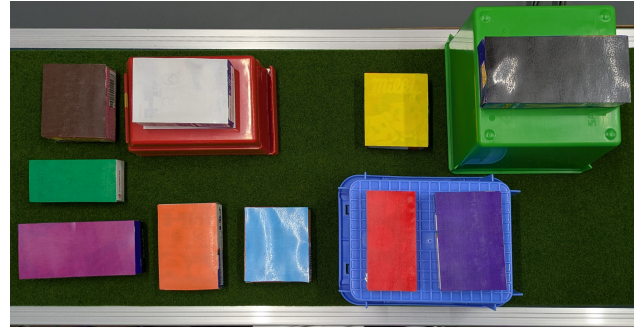


図 7 本番 2 の配置環境

表 4 実験セッションの構成

セッション	配置数	把持数	目的
練習 1	2	1	操作の習熟
練習 2	3	1	操作の習熟
本番 1	5	3	中難易度環境での評価
本番 2	10	5	高難易度環境での評価

箱を配置した高難易度環境とした。

4.5 実験手順

実験は以下の手順で実施した。

- (1) 被験者に実験の目的・手順・安全上の注意を説明し、書面による同意を取得
- (2) HMD を装着し、割り当てられた手法の操作方法を説明
- (3) 練習 1・練習 2 を実施し、操作への十分な習熟を確認
- (4) 本番 1・本番 2 を実施
- (5) HMD を取り外し、アンケートに回答

本番セッションでは、各把持試行のタスク完遂時間を計測した。

4.6 評価指標

以下の 4 種類の指標を用いて各手法を評価した。

- **タスク完遂時間**：各把持試行における初期姿勢から自動把持を開始するまでの時間（秒）
- **SUS スコア**：10 項目・100 点満点のシステム全体の主観的使いやすさ評価 [21]
- **NASA-TLX (NASA Task Load Index) スコア**：精神的負荷・身体的負荷・時間的負荷・作業達成度・努力・不満の 6 下位尺度から構成される主観的作業負荷評価
- **信頼感**：自動把持機能に対する操作者の信頼感をリッカート（Likert）尺度で評価
- **自由記述**：操作感や改善要望に関する定性的なフィードバックの収集

4.7 実験結果

表 5 に各指標の結果をまとめる。3 手法の比較には

Kruskal-Wallis 検定を用いた。

4.7.1 タスク完遂時間

図 8 に各手法のタスク完遂時間を示す。

本番 1 では、手先位置選択が 9.6 ± 6.0 s、最近傍選択が 11.8 ± 6.6 s、スティック選択が 35.1 ± 30.3 s となり、手先位置選択が最短時間を示した。本番 2 では、手先位置選択が 48.6 ± 19.7 s、最近傍選択が 54.7 ± 25.4 s、スティック選択が 60.1 ± 15.0 s となり、本番 1 と同様の傾向が確認された。いずれのセッションでも 3 手法間に統計的有意差は認められなかった（本番 1: $H = 4.308$, $p = 0.116$; 本番 2: $H = 0.269$, $p = 0.874$ ）。

4.7.2 SUS スコア・NASA-TLX・信頼感

図 9 に各手法の主観評価結果を示す。

SUS スコアは手先位置選択が 61.2 ± 6.0 、最近傍選択が 56.2 ± 12.5 、スティック選択が 61.2 ± 6.6 であり、最近傍選択がやや低い値を示した（ $H = 0.363$, $p = 0.834$ ）。NASA-TLX スコアは手先位置選択が 3.3 ± 0.7 、最近傍選択が 4.4 ± 2.1 、スティック選択が 3.4 ± 1.4 であり、最近傍選択がやや高い値を示した（ $H = 1.081$, $p = 0.583$ ）。信頼感は手先位置選択が 4.0 ± 0.3 、最近傍選択が 3.9 ± 0.4 、スティック選択が 4.1 ± 0.4 と 3 手法でほぼ同水準であった（ $H = 1.015$, $p = 0.602$ ）。いずれの指標でも統計的有意差は認められなかった。

4.7.3 定性的フィードバック

自由記述において得られた主なコメントを手法別に整理する。各手法に共通する課題として、奥行き知覚の補助とモード状態の可視化が挙げられた。

手先位置選択の使用者からは、「立体的に空間を把握するのが難しかった」、「仮想アームから下に向けてビームを描写することで奥行きが計りやすくなるかもしれない」といった、VR 映像での奥行き把握に関するコメントが得られた。最近傍選択の使用者からは、「アームが動いているのか分からず A ボタンが反応したか判断しづらかった」、「自由移動中か自動操作かわかるとありがたい」といった、モード状態の可視化に関する要望が得られた。スティック選択の使用者からは、「A ボタンを押す瞬間にターゲットが変わったりすることがあった」、「アームの移動が必要なときとスティックで選択の方が早い時があると思ったから、その補助があるとおもしろいと思った」といった、操作タイミングの安定性や複合的な支援に関するコメントが得られた。

5. 考察

5.1 手法間の比較

本実験では、手先位置選択が完遂時間・NASA-TLX の観点で 3 手法中最も安定した値を示す傾向が確認された。第 4.1 節の予備実験で述べた通り、最近傍選択は手先位置選択の上位互換として設計しており、平面的な対象物を扱っ

た環境では同等以上の性能が示されていたが、三次元的に複雑な配置を用いた本実験ではその傾向が逆転し支持されなかった。

「自由移動中か自動操作かわかるとありがたい」という自由記述が示すように、A ボタン押下後にシステムが自動的に最近傍の物体を選択するため、VR 映像の奥行き知覚制約により、高さの異なる物体が混在する本実験環境では、操作者が意図した物体と最近傍として判定される物体が一致しない場面が生じやすかったと考えられる。予備実験のような単純な環境では意図しない選択が生じるリスクが低かったが、本実験ではこのシステム挙動の不透明性が、操作の簡略化という恩恵を上回り操作感の低下につながった可能性がある。

スティック選択は本番 1 の完遂時間が 35.1 ± 30.3 s と最も長く、1 名が 79.27 s を要するなど個人差が大きかった。「A ボタンを押す瞬間にターゲットが変わった」という指摘が示すように、スティック操作中の候補切り替えが不安定な場面があったと考えられる。

なお、全指標で統計的有意差は認められなかった。各群 4 名という被験者数の少なさから統計的検出力が低く、現時点での差異は傾向の把握に留まる。

5.2 配置密度による傾向の変化

本番 1 と本番 2 の比較から、配置密度が完遂時間に与える影響を考察する。全 3 手法で本番 2 の完遂時間は本番 1 を大きく上回り、配置密度の増加が操作時間の増大につながる傾向が確認された。手先位置選択は本番 1 から本番 2 で完遂時間が約 5 倍に増加したのに対し、スティック選択は約 1.7 倍の増加に留まり、高密度環境でのスティック選択の相対的な安定性が示唆された。これは、スティック選択では対象物への精密な位置操作が不要であり、配置密度が増加しても操作手順が変化しないためと考えられる。

5.3 明示的切替型構造に対する評価

信頼感スコアは 3 手法いずれも 4.0 前後と高水準であり、自動把持機能への信頼感は手法によらず一定程度確保された。自由記述においても、操作状態の可視化や奥行き知覚の補助に関する改善要望が得られたものの、システムへの根本的な不満を示すコメントは少なく、明示的切替型構造が操作者に受け入れられやすい可能性が示唆された。

一方で、「自由移動中か自動操作かわかるとありがたい」という指摘に代表されるように、モード状態の可視化が課題として浮かび上がった。明示的切替型では操作者がモードを制御するため透明性は高いが、自動シーケンス中のロボット状態を VR 映像上でより明確に提示すれば、操作感のさらなる向上につながると考えられる。

表 5 各手法の実験結果（平均 ± 標準偏差）

手法	完遂時間 本番 1 (s)	完遂時間 本番 2 (s)	SUS スコア	NASA-TLX	信頼感
手先位置選択	9.6 ± 6.0	48.6 ± 19.7	61.2 ± 6.0	3.3 ± 0.7	4.0 ± 0.3
最近傍選択	11.8 ± 6.6	54.7 ± 25.4	56.2 ± 12.5	4.4 ± 2.1	3.9 ± 0.4
スティック選択	35.1 ± 30.3	60.1 ± 15.0	61.2 ± 6.6	3.4 ± 1.4	4.1 ± 0.4
p 値 (Kruskal-Wallis)	0.116	0.874	0.834	0.583	0.602

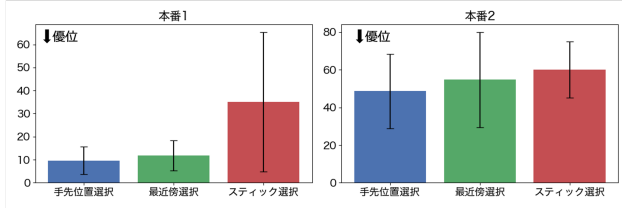


図 8 手法別タスク完遂時間の比較

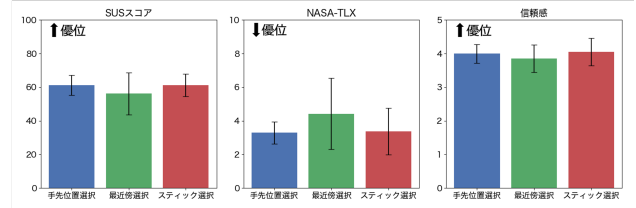


図 9 手法別 SUS スコア・NASA-TLX・信頼感の比較

5.4 限界と今後の課題

本実験にはいくつかの限界が存在する。第一に、各手法 4 名という少数の被験者数から統計的検出力が低く、手法間の差異は傾向の把握に留まる。第二に、把持対象が単一種類の箱であり、形状が多様な実環境の対象物を網羅していない。第三に、テーブル上の静的な配置のみを対象としており、動的に変化する環境での評価は行っていない。第四に、本実験の設計は選択手法間の差異を明確に分離するには不十分であった可能性がある。タスク完遂時間には位置操作と選択操作の両フェーズが含まれており、手法固有の特性よりも操作者の位置操作スキルや UI への適応が結果に影響した可能性がある。これらを踏まえ、今後は被験者数の拡大、多様な対象物・環境条件での評価、モード状態の可視化機能の実装と評価に加え、位置操作フェーズを統制した評価設計や手法ごとに適した UI フィードバックの導入による追加実験に取り組む。

6. 結論

本研究では、VR 映像を用いた遠隔ロボットアーム操作システムにおいて、操作者主導の明示的切替型自動把持支援を設計し、実環境で動作する Dobot Nova 2 に実装した。位置操作から選択操作への移行タイミングが異なる 3 種類の対象物選択手法（手先位置選択・最近傍選択・スティック選択）を、連続的な設計空間の代表点として実装し、12 名の被験者による把持タスク実験を通じて比較評価を行った。

実験の結果、3 手法間に統計的有意差は認められなかった。完遂時間や操作負荷の観点では手先位置選択が最も安定した値を示したが、これは平面的な配置を用いた予備実験において最近傍選択が最も良好な成績を示した結果とは逆転する形となった。各群 4 名という少数の被験者数に加え、選択手法の差異を分離して評価するには実験設計上の改善が必要であり、手法間の性能差についての結論は追加実験による検証が必要である。高密度環境ではスティック選択が完遂時間の増加率が低く、密度条件による手法の適

性の違いが示唆された。

今後の課題として、まず被験者数の拡大に加え、位置操作フェーズと選択フェーズを分離した評価設計や手法ごとに適した UI フィードバックを導入した上での比較実験により、各選択手法の特性をより正確に検証する必要がある。また、形状が多様な対象物や動的に変化する環境での評価を行い、各手法の汎用性と密度条件に応じた適性を明らかにする。さらに、自動シーケンス中のモード状態の可視化機能の実装と、深度推定モデルを用いた三次元位置推定の高精度化にも取り組む。

謝辞 本研究の一部は、内閣府総合科学技術・イノベーション会議の「SIP/バーチャルエコノミー拡大に向けた基盤技術・ルールの整備」(JPJ012495) (研究推進法人：国立研究開発法人新エネルギー・産業技術総合開発機構) によって実施されました。

参考文献

- [1] Darvish, K., Penco, L., Ramos, J., Cisneros-Limón, R., Pratt, J., Yoshida, E., Ivaldi, S. and Pucci, D.: Teleoperation of Humanoid Robots: A Survey, *IEEE Transactions on Robotics*, Vol. 39, No. 3, pp. 1706–1727 (2023).
- [2] Lipton, J. I., Fay, A. J. and Rus, D.: Baxter's Homunculus: Virtual Reality Spaces for Teleoperation in Manufacturing, *IEEE Robotics and Automation Letters*, Vol. 3, No. 1, pp. 179–186 (2018).
- [3] Wonsick, M. and Padir, T.: A Systematic Review of Virtual Reality Interfaces for Controlling and Interacting with Robots, *Applied Sciences*, Vol. 10, No. 24, p. 9051 (2020).
- [4] Vienne, C., Sorin, L., Blondé, L., Huynh-Thu, Q. and Mamassian, P.: Depth Perception in Virtual Reality Systems: Effect of Screen Distance, Environment Richness and Display Factors, *IEEE Access*, Vol. 8, pp. 26414–26422 (2020).
- [5] Dragan, A. D. and Srinivasa, S.: A Policy-Blending Formalism for Shared Control, *The International Journal of Robotics Research*, Vol. 32, No. 7, pp. 790–805 (2013).
- [6] Atan, U., Bharadwaj, V. R. and Jiang, C.: Assistive Control of Robot Arms via Adaptive Shared Autonomy, *Proc. IEEE/ASME AIM* (2024).

- [7] Goertz, R. C. et al.: General Purpose Manipulators, *Nucleonics*, Vol. 12, No. 11, pp. 45–49 (1954).
- [8] Hokayem, P. F. and Spong, M. W.: Bilateral Teleoperation: An Historical Survey, *Automatica*, Vol. 42, No. 12, pp. 2035–2057 (2006).
- [9] Whitney, D., Rosen, E., Ullman, D., Phillips, E. and Tellex, S.: ROS Reality: A Virtual Reality Framework Using Consumer-Grade Hardware for ROS-Enabled Robots, *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, pp. 1–9 (2018).
- [10] Stotko, P., Krumpen, S., Hullin, M. B., Weinmann, M. and Klein, R.: A VR System for Immersive Teleoperation and Live Exploration with a Mobile Robot, *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, pp. 3630–3637 (2019).
- [11] Arevalo Arboleda, S., Rücker, F., Dierks, T. and Gerken, J.: Assisting Manipulation and Grasping in Robot Teleoperation with Augmented Reality Visual Cues, *Proc. CHI Conf. Human Factors Comput. Syst.*, pp. 1–14 (2021).
- [12] Zhu, Y., Jiang, B., Chen, Q., Aoyama, T. and Hasegawa, Y.: A Shared Control Framework for Enhanced Grasping Performance in Teleoperation, *IEEE Access* (2023).
- [13] Javdani, S., Admoni, H., Pellegrinelli, S., Srinivasa, S. S. and Bagnell, J. A.: Shared Autonomy via Hindsight Optimization for Teleoperation and Teaming, *The International Journal of Robotics Research*, Vol. 37, No. 7, pp. 717–742 (2018).
- [14] Reddy, S., Dragan, A. D. and Levine, S.: Shared Autonomy via Deep Reinforcement Learning, *Proc. Robotics: Sci. Syst. (RSS)* (2018).
- [15] Kleeberger, K., Bormann, R., Kraus, W. and Huber, M. F.: A Survey on Learning-Based Robotic Grasping, *Current Robotics Reports*, Vol. 1, pp. 239–249 (2020).
- [16] Newbury, R., Gu, M., Chumbley, L., Mousavian, A., Eppner, C., Leitner, J., Bohg, J., Morales, A., Asfour, T., Kragic, D., Fox, D. and Cosgun, A.: Deep Learning Approaches to Grasp Synthesis: A Review, *IEEE Transactions on Robotics*, Vol. 39, No. 5, pp. 3994–4015 (2023).
- [17] Tsutsui, S., Watanabe, K., Usami, R., Gushi, Y., Watanabe, Y., Shimura, K., Hayashida, N., Jimenez, F., Urano, K., Yonezawa, T. and Kawaguchi, N.: Mitigating Latency Effects in VR180 Teleoperation via Virtual Robotic Arm Overlay, *Proc. ICMU* (2025).
- [18] Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W. and Wang, X.: ByteTrack: Multi-Object Tracking by Associating Every Detection Box, *Proc. ECCV*, pp. 1–21 (2022).
- [19] Zhao, X., Ding, W., An, Y., Du, Y., Yu, T., Li, M., Tang, M. and Wang, J.: Fast Segment Anything, *arXiv:2306.12156* (2023).
- [20] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P. and Girshick, R.: Segment Anything, *Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV)*, pp. 3992–4003 (2023).
- [21] Brooke, J.: SUS: A “Quick and Dirty” Usability Scale, *Usability Evaluation In Industry*, pp. 189–194 (1996).