モバイルマニピュレータによる動きながらのPick-and-Placeを可能にする認識システムの開発

◯小野 智寛　竹下 佳佑　山崎 隆広　新谷 和宏（トヨタ自動車株式会社）

近年，モバイルマニピュレータでのPick-and-Placeは一般的な物体操作において，高度なレベルに達しているが，未だStop-and-Goでの動作が主流である．動きながらのPick-and-Placeを実現するためには，リアルタイム処理が必要不可欠であり，さらに高い精度が要求される．本研究では，スムーズな動作を実現するための，一連の認識システムを提案する．提案システムをTOYOTA HSRに搭載して評価し，従来手法と比較して1.4倍の速さで片付けタスクを実行できることを確認した．

1.　はじめに

近年，少子高齢化社会による人手不足が深刻な問題となっている．この問題を解決するため，我々は安全かつ小型で，動的環境下で人間と協働して作業を行う生活支援ロボットHuman Support Robot（HSR）の研究開発を進めている[1]．我々はHSRを活用し，人間との協働作業において，速さ（Speed），滑らかさ（Smooth/Smart），安定（Stable），安全（Safe）という4Sの実現を目指している．

これまでに，動作間で逐次停止をする，Stop-and-GoでのPick-and-Placeタスクは高度なレベルで実現されてきている[2, 3]．一方で，一切停止をせず，動きながらのPick-and-Placeタスクでは，多様な種類の物体に対応できていない，環境にカメラの設置が必要であるなど，課題が多いのが現状である[4, 5]．

また，モバイルマニピュレータでの人協働作業において，4Sをテーマに開かれているWorld Robot Summit（WRS）では，HSRを用いてPick-and-Placeタスクの代表例である片付けタスクを行い，その達成度を競う競技がある． 15分間の競技で，使用される31個すべての物体を片付けるためには，1つの物体をおよそ30秒以内で片付ける必要がある． 2021年に開かれた競技会では，9チームが参加したが，どのチームもStop-and-Goで動作しており，16個の片付け（約52%）が最高であった[6]．

我々は，HSRを用いて，リアクティブな振る舞いが可能な，周期的な全身軌道計画手法を提案してきた[7, 8]．この手法に，高速かつ頑健な認識を組み合わせることで，4Sを実現するソフトウェアを確立し，動きながらのPick-and-Placeタスクを実現することを目指す．

本研究では，HSRに搭載されているセンサのみを用いた，Pick-and-Placeタスクのための認識システムを提案する．ヘッドカメラによるラフな物体位置推定と記憶，ハンドカメラによる正確な把持姿勢推定，そして，これらのカメラ間での同一物体トラッキングを組み合わせることで，動作を止めることなく認識を行い，タスクを実行する．実験では，複数の物体を片付ける際の，PickおよびPlaceの成功率と実行時間を従来手法と比較し，有効性を検証する．

2.　提案手法

2.1　概要

図1に提案手法の概要を示す．提案するシステムは，ヘッドカメラによる物体認識，3次元位置算出および矩形近似，ハンドカメラによる物体追跡および把持姿勢推定で構成されている．また，2つのカメラ間は，記憶モジュールを介して同一物体の追跡を実現している．なお，ヘッドカメラは標準搭載のXtion PRO LIVE，ハンドカメラには追加で搭載したOrbbec Gemini2を使用している．

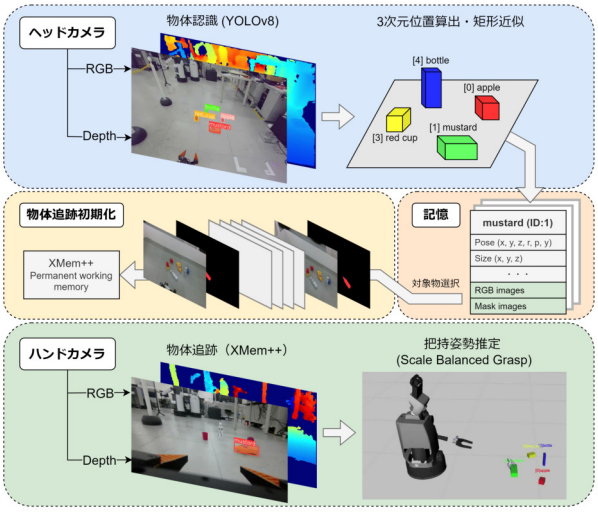

図1　提案手法の概要．

2.2　物体認識

ヘッドカメラでの物体認識では，台車や頭部の動作により映像に大きな変化が生じるため，リアルタイム処理が必須である．そこで，リアルタイムでInstance Segmentationを実行できるYOLOv8[9]を用いている．学習データの生成には，IsaacSimを用いており，カメラの移動によって生じるブラーに対して頑健になるようにデータを大量生成し，Sim2Realでの認識を実現している．また，認識したマスク画像とDepth画像を用いて，各物体の3次元点群を生成し，3次元の位置算出および矩形近似を行う．これらの情報は，次に述べる記憶モジュールで使用する．

2.3　記憶モジュール

記憶モジュールでは，物体認識によって得られた情報（時刻，クラス，座標，点群，RGB画像，マスク画像，Bounding Box）を追加，保持，更新，忘却する機能を持つ．記憶モジュールの手順は以下の通りである．

まず，過去から現在までに観測された情報に対して，各物体ごとに3次元点群の距離であるChamfer Distance

図2　ロバストな把持姿勢の抽出.

（CD）を次式により計算する.

$$d_{cd}(X,Y) = \frac{1}{|X|}\sum_{x\in X}\min_{y\in Y}\|x-y\|_2^2 + \frac{1}{|Y|}\sum_{y\in Y}\min_{x\in X}\|x-y\|_2^2 \quad (1)$$

ここで，XおよびYは物体の点群，xおよびyはそれぞれ，点群Xおよび点群Yのある一点である.

次に，計算されたCDを用いて，HDBSCAN[10]によりクラスタリングを行う．クラスタリングされた点群ごとに，過去の記憶された物体情報との，3次元矩形のIntersection over Union（IoU）を計算する． IoUの重なりがしきい値以下であれば記憶への新規追加，しきい値以上であれば情報の更新を行う．また，記憶されている物体情報をもとに，同じ位置で一定時間観測されていない場合，忘却を行う.

2.4　物体追跡

物体追跡では，XMem++[11]を用いて，ヘッドカメラとハンドカメラで同一物体の追跡を実現する．タスク実行の際には，記憶モジュールからロボットに最も近い物体のRGBおよびマスク画像を複数枚取得する．それらの画像を用いて，XMem++のPermanent working memoryを初期化し，ハンドカメラで追跡を行う.

2.5　把持姿勢推定

物体追跡により得たセグメンテーションを用いて，Scale Balanced Grasp（SBG）[12]により物体の把持姿勢を推定する． SBGはTabletopのデータを用いてTop-downで把持できるように学習されているため，モバイルマニピュレータにとって適切でない把持候補が出力されることがある.

そこで，本研究では，図2に示すように，スコアリングの修正とロバストな把持姿勢の抽出を行っている．図2(a)はSBGが出力する把持候補を可視化したものである．候補が赤ければ赤いほどスコアが高いことを表す.

図2(b)はスコアリング修正後の把持候補を可視化したものである．ここでは，物体中心を把持したほうが安定すると仮定し，各把持候補と物体との接点を求め，その点が物体の中心に近ければ近いほどスコアを高くする．図2(b)を見ると，物体中心に近いほど候補が赤くなっていることがわかる.

図2(c)は，ロバストな把持姿勢のみを可視化したものである．ここでは，まず，全把持候補に対して式(2)から式(4)に示す，Grasp Non-Maximum Suppression（Grasp NMS）を行い，近傍の把持候補をまとめてG_Cとする．次に，並進および回転誤差の許容度

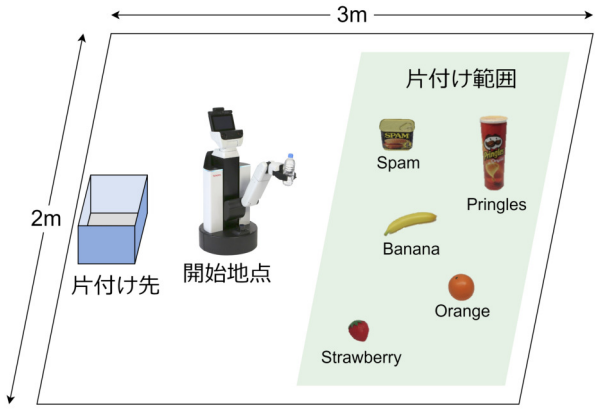

図3　実験に使用した環境および物体.

を表すRを式(5)を用いて計算する．最後に，Rがしきい値を超えた把持候補のみを使用する．図2(c)はしきい値を2.0にしたときの把持候補を示している.

$$d_t(G_m,G_n) = \|\mathbf{t_1}-\mathbf{t_2}\| \quad (2)$$

$$d_\alpha(G_m,G_n) = \cos^{-1}\left(\frac{1}{2}(tr(R_1\cdot R_2^T)-1)\right) \quad (3)$$

$$G_C = \{(G_m,G_n)\mid d_t(G_m,G_n)<th_d \wedge\ d_\alpha(G_m,G_n)<th_\alpha\} \quad (4)$$

$$R(G_C) = \sum_{i=1}^{N}\left(\frac{(d_t(G_c,G_i)}{th_d}+\frac{(d_\alpha(G_c,G_i)}{th_\alpha}\right) \quad (5)$$

ここで，G_m，G_nは把持候補，d_tは並進誤差，d_αは回転誤差，$\mathbf{t}$は並進ベクトル，Rは回転行列，$tr(M)$は行列Mのトレース，th_dは並進誤差のしきい値，th_αは回転誤差のしきい値，NはGrasp NMS後のG_Cに含まれる把持候補数を表す.

3.　実験

3.1　概要

本手法の有効性を確かめるために，WRSタスクを簡易的に模擬した環境で片付けタスクを行った．実験環境は，図3に示す通りである．物体は，YCBオブジェクト[13]から大きさ，形状の異なる5種類の物体を選んで使用し，片付け範囲内に適当に散らばせた状態からタスクを開始した．従来手法[6]および提案手法でそれぞれタスクを30回の実行し，タスク実行に要した時間，PickおよびPlaceの成功率を計測した． Pickは，対象物を把持し，持ち上げることができたら成功と判断する．また，Pick失敗後に自律的に再試行し，成功した場合は，成功とみなす． Placeは，Pick成功後に片付け先まで落とさずに移動し，箱の中に入れることができれば成功とする.

3.2　結果

実験結果を1に示す．提案手法では，従来手法と比較して約1.4倍の速度で片付けタスクを完了させることができた．また，動きながら認識を行った場合でも，PickおよびPlaceの成功率を低下させることはなく，むしろStop-and-Goである従来手法よりも，Pick成功率

	従来手法	提案手法
実行時間 [s]	122.0 ± 5.87	**87.56** ± 8.4
Pick成功率 [%]	94.0 (141/150)	**96.0** (144/150)
Place成功率 [%]	99.29 (140/141)	**99.31** (143/144)

表1　実験結果.

図4　従来手法と提案手法の把持姿勢の違い.

は2ポイント高い結果となった．Placeの失敗は，どちらも手法も1回のみであり，移動中に落下しているため，実質的には把持姿勢が悪いと考えられる．

図4に従来手法と提案手法の把持姿勢の違いを示す．従来手法では，ヘッドカメラを用いて物体の点群を取得し，主成分分析で物体の回転方向を算出，上から把持するのが基本戦略である．ヘッドカメラから物体までの距離は遠くなるため，Depth精度が低下しやすくなる影響で，把持姿勢が良くない傾向があった．また，基本的に上から掴むため，物体によっては安定した把持姿勢でないことが多く，精度の低下につながったと考えられる．一方，提案手法では，ハンドカメラを用いて物体の点群を取得するため，物体との距離が近く，Depth精度が低下しづらい．また，物体の形状を考慮してロバストな把持姿勢を算出しているため，安定した把持姿勢を算出できていた．

4.　おわりに

本研究では，4Sをテーマに，動きながらPick-and-Placeタスクを実行するための，認識システムを提案した．提案手法は，従来手法と比較して，把持精度を向上させ，約1.4倍の速度で片付けタスクを実行することができた．

提案手法では，1物体あたり約18秒で片付けが完了しており，WRSタスクに適用した場合，約9.2分で31個全ての物体を片付けることが見込まれる．一方で，今回行った実験は物体の種類が少ない，環境が簡単という条件であったため，そのままWRSタスクに適用することは難しい．また，4Sのうちの「安全」には考慮できておらず，物体を落とさない（優しく置く），環境を壊さないなどにも取り組む必要がある．今後は，速さ，滑らかさ，安定を維持しつつ，さらに安全にも考慮した上で，WRSタスクの完遂を目指して，さらなる研究開発を行っていく．

参　考　文　献

[1] T. Yamamoto, K. Terada, A. Ochiai, F. Saito, Y. Asahara, and K. Murase. Development of human support robot as the research platform of a domestic mobile manipulator. *ROBOMECH Journal*, 6(1), 12 2019.

[2] J. Wu, R. Antonova, A. Kan, M. Lepert, A. Zeng, S. Song, J. Bohg, S. Rusinkiewicz, and T. Funkhouser. Tidybot: Personalized robot assistance with large language models. *Autonomous Robots*, 2023.

[3] P. Liu, Y. Orru, C. Paxton, N. M. M. Shafiullah, and L. Pinto. Ok-robot: What really matters in integrating open-knowledge models for robotics. *arXiv:2401.12202*, 2024.

[4] B. Burgess-Limerick, C. Lehnert, J. Leitner, and P. Corke. An architecture for reactive mobile manipulation on-the-move. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1623–1629, 2023.

[5] 長田瑛綺, 清川拓哉, 鶴峯義久, 権裕煥, and 松原崇充. 視覚情報に基づく移動把持の自己教師あり学習. ロボティクス・メカトロニクス講演会講演概要集, 2024:1P1–H04, 5 2024.

[6] T. Ono, D. Kanaoka, T. Shiba, Y. Tokuno, Shoshi abd Yano, A. Mizutani, I. Matsumoto, H. Amano, and H. Tamukoh. Solution of world robot challenge 2020 partner robot challenge (real space). *Advanced Robotics*, 36(17-18):870–889, 2022.

[7] 竹下佳佑 and 山本貴史. 周期的な全身軌道計画を用いたmobile manipulation システムの提案. 第*28*回ロボティクスシンポジア, 3 2023.

[8] 竹下佳祐 and 山本貴史. 生活支援ロボットhsrの台車位置をパラメータとした全身ik. 第*41*回日本ロボット学会学術講演会予稿集, 1J2-02, 9 2023.

[9] G. Jocher, A. Chaurasia, and J. Qiu. Ultralytics YOLO, January 2023.

[10] L. McInnes, J. Healy, and S. Astels. hdbscan: Hierarchical density based clustering. *Journal of Open Source Software*, 2(11):205, 2017.

[11] M. Bekuzarov, A. Bermudez, J.-Y. Lee, and H. Li. Xmem++: Production-level video segmentation from few annotated frames. *arXiv:2307.15958*, 2023.

[12] M. Haoxiang and D. Huang. Towards scale balanced 6-dof grasp detection in cluttered scenes. In *Conference on Robot Learning (CoRL)*, 2022.

[13] B. Calli, A. Singh, A. Walsman, S. Srinivasa, P. Abbeel, and A. Dollar. The ycb object and model set: Towards common benchmarks for manipulation research. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 510–517, 07 2015.