

ホームロボットによる物体姿勢推定のための 仮想空間での推論に基づく視点計画

○原田悠斗（大阪大） 青木達哉（大阪大）
堀井隆斗（大阪大） 長井隆行（大阪大／電気通信大）

1. はじめに

人々の生活空間へと活動領域を拡大させているホームロボットが、人間の要求に応じて適切にタスクを達成するには環境を正しく認識する必要がある。例えば机上の物を掴み、所定の位置に片づけるタスクでは、物体の位置と姿勢を正確に推定する能力が重要である。近年の深層学習の技術の発展に伴い、PoseCNN [1] や DenseFusion [2] といった物体認識や姿勢推定モデルは、高い推定精度を発揮する。しかし、これらのモデルを実ロボットで利用する際には、その推定精度が大きく低下する場合がある。これは、物体の姿勢推定モデルが未学習の物体姿勢や、姿勢を一意に決定することが困難な物体姿勢など、推定誤差が大きくなる位置からロボットが物体を観測していることが原因として挙げられる。特に、ロボットと対象物体の位置が変動するような把持タスクでは、これらの問題が頻出すると予想される。

ロボットにおける物体認識の低下を防ぐ手法として、図 1 に示すように物体の観測位置を適切に選択する視点計画が提案されている [3,4]。McGreavy ら [3] は、オンラインのアスペクトグラフを用いて新しい視点での対象物体の可視性を分析し、オクルージョンを考慮した視点計画手法を提案した。また Zhang ら [4] は、物体と場所や物体同士の位置関係に関する事前知識に基づく物体探索のための移動戦略を提案した。しかし、これらの研究は物体認識のための視点計画にとどまり、姿勢推定精度の向上については言及されていない。

本研究では、ロボットが物体を認識できる観測位置から、実空間情報をレンダリングした仮想空間上での物体姿勢の推論を通じて、対象物体の姿勢を正確に推定するための観測位置を決定する視点計画手法を提案する。本稿では、提案手法による物体姿勢推定精度の変化をシミュレーションと実機実験により検証した。また、実ロボットによる把持タスクによって提案手法の有用性を示した。

2. 提案手法

物体の姿勢推定精度を向上させる視点計画を効率的に実現するためには、実空間におけるロボットの移動コスト削減と、真値が不明な物体姿勢の推定精度を評価する指標が必要となる。本研究ではこれらの要求を満たすために、実空間に存在する対象物体をロボットの持つ仮想空間にレンダリングし、仮想空間中の複数の視点から物体姿勢を推論することで、効率的な視点探索を実現する。また、ロボットが観測している実空間中の物体画像と、現在の姿勢推定結果に基づいて仮想空間にレンダリングされた対象物体の画像の類似度

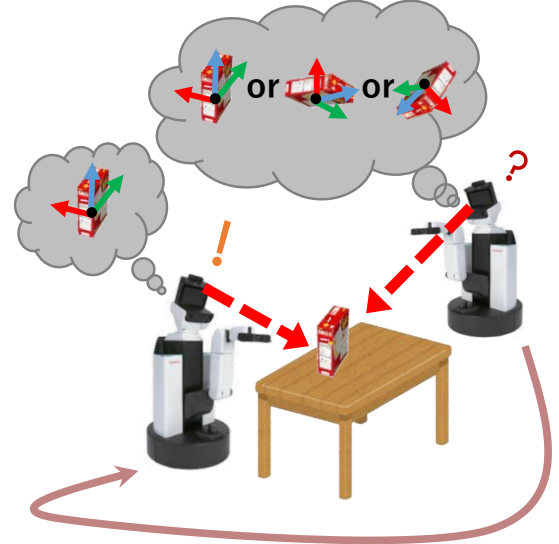


図 1 ロボットによる姿勢推定結果の評価と視点計画

から、姿勢推定結果の正確性を間接的に評価する。

上記の機能を導入した視点計画手法の流れを図 2 に示す。提案手法は、現在の観測位置における姿勢推定結果の正確性評価 ((a)~(d)) と、新しい観測位置選択 ((e) と (f)) の 2 つの処理で構成される。以降、それぞれの処理を各 step ごとに説明する。

2.1 推定姿勢の正確性の評価

はじめに、図 2 中の step (a)~(d) について述べる。これらの step では、実空間で対象物体の姿勢を推定し、観測画像の類似度を用いて推定姿勢の正確性を定量的に評価する。その後、現在の観測位置が適切であるかを判定する。

Step(a)：姿勢推定

同じ観測位置においても外乱によって姿勢推定の結果にばらつきが生じる。そこで、1 つの観測位置で複数枚の物体画像を取得し、それぞれに対して姿勢推定を実施する。

Step(b)：姿勢 P_{now} の選択

Step (a) で推定した各姿勢から、推定誤差が最小と思われる姿勢 P_{now} を選択する。まず、実空間での姿勢推定結果に基づき、対象物体を仮想空間上でレンダリングする。その後、仮想空間上で実世界と同じ観測位置から画像 I_S を取得し、実空間の画像 I_R との類似度を領域ベースマッチングにより算出する。本研究では

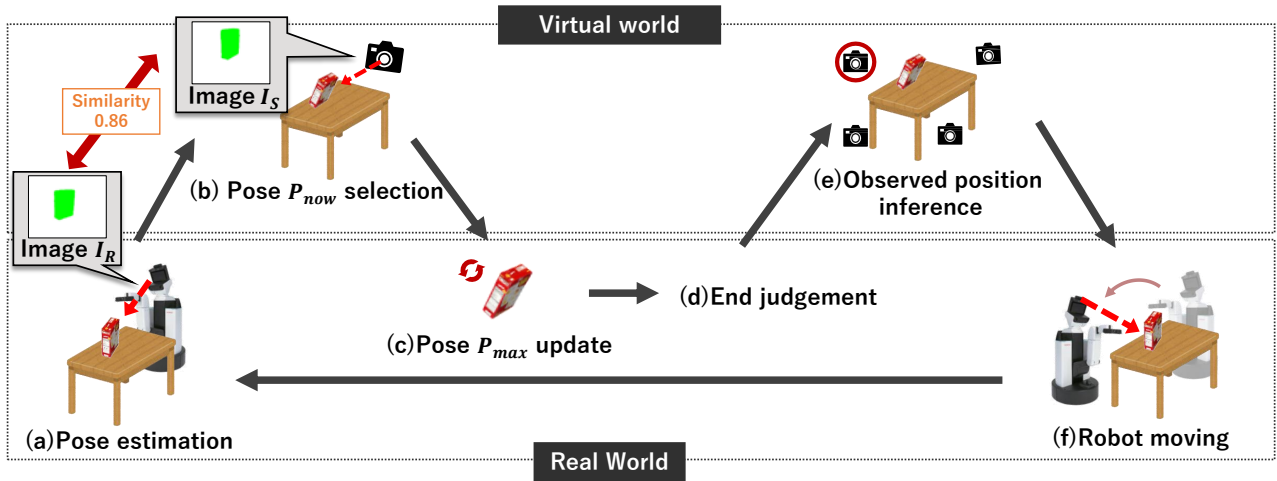


図 2 提案手法の流れ

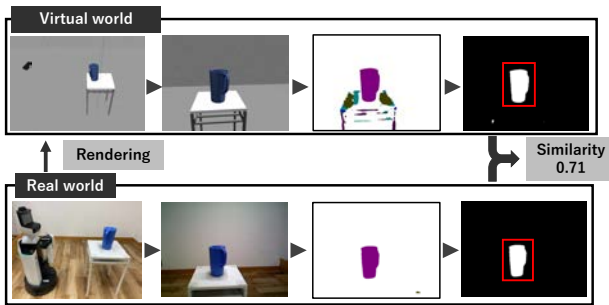


図 3 画像の類似度の計算

物体姿勢のみに着目するため、図 3 に示すように認識物体ごとに色付けした画像に対して二値化処理を施し、対象物体領域を抽出した画像から類似度を求める。複数枚取得した物体画像の各姿勢推定結果に対して類似度を計算し、最大の類似度を有する姿勢 P_{now} を選択する。

Step(c)：姿勢 P_{max} の更

視点計画開始時から現在までの姿勢推定において類似度が最大の姿勢 P_{max} と現在の観測位置での類似度が最大の姿勢推定結果 P_{now} を比較し、類似度が大きいもので P_{max} を更新する。

Step(d)：終了判定

現在の観測位置が適切であるかを以下の 2 つの観点から判定し、両方の条件を満たす場合のみ視点計画手法を終了する。

正確性条件

現在の観測位置で取得した姿勢推定結果 P_{now} の誤差が十分小さいと見込まれるかを判定する。姿勢推定結果 P_{now} の画像の類似度が設定した閾値以上であることを条件とし、その閾値は経験的に決定する。

再現性条件

Step (a) で得られた複数の姿勢推定結果が、現在

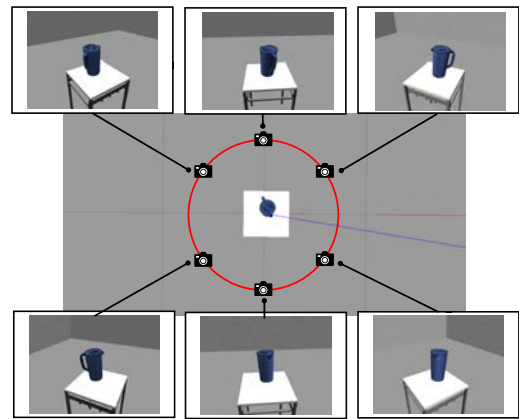


図 4 仮想空間での姿勢推定

の観測位置で得られた最良の姿勢推定結果 P_{now} に対してどれだけ近いかを評価する。姿勢 P_{now} に対する各推定結果の平均誤差が閾値以下であることを条件とする。これにより、step (a) で述べた外乱による姿勢推定のばらつきが生じにくい観測位置を選択できる。本研究では、姿勢間の距離計算に式 (1) で定義される ADD-S [1] を用いる。ここで比較する各姿勢の回転行列と平行移動行列を (R, T) , (\tilde{R}, \tilde{T}) とし、対象物体の 3D モデル上からサンプリングされた m 個の点群を \mathcal{M} とする。

$$\frac{1}{m} \sum_{x_1 \in \mathcal{M}} \min_{x_2 \in \mathcal{M}} \|(Rx_1 + T) - (\tilde{R}x_2 + \tilde{T})\| \quad (1)$$

サンプリング数は 100 点とし、実験で用いる物体の 3D モデルの点群を使用した。

終了条件を満たす観測位置が存在しない場合、観測点の探索が無限に続くことになる。これを防ぐために、移動後の観測位置において step (c) の更新が行われず、かつ移動後の姿勢推定結果の平均誤差が step (d) の推論時に想定した推定誤差未満である場合に視点計画を終了する。



図5 YCB オブジェクト



図6 実験環境: (a) シミュレーション環境, (b) 実機環境

2.2 観測位置の選択

終了条件を満たさない場合、推定誤差が減少する観測位置を仮想空間内で探索する。これにより未観測位置での姿勢推定誤差を予測し、ロボットの効率的な視点選択を実現する。

Step (e)：多視点での姿勢推定

物体を姿勢 P_{\max} で仮想空間へレンダリングした後、図4の様に物体を中心とした円周上に 60° おきにカメラを配置し、物体姿勢を推定する。カメラは、床に平行かつ対象物体が中央に移る姿勢で、現在のロボットカメラの高さに揃えて配置する。また、カメラの高さによる姿勢推定精度の変化についても考慮するため、step (c) で姿勢 P_{\max} が更新されていない場合は配置するロボットカメラの高さを変更する。

Step (f)：観測位置の選択とロボットの移動

姿勢 P_{\max} に対する仮想空間上の各観測位置での姿勢推定結果の誤差を ADD-S により計算する。その後、誤差が最小の観測位置に実空間のロボットを移動させる。

3. 実験

シミュレーションと実機実験により提案手法の有用性を検証した。実験環境を図6に示す。環境中には対象物体を設置するテーブルのみが存在する。また両実験共に、姿勢推定対象物体以外の環境物体の位置情報は既知という前提で仮想空間を構築し、実空間での推定姿勢に基づいて対象物体のみをレンダリングする。この際、レンダリング時に重力の影響で物体が転倒して姿勢が変化することを防ぐために、重力に関する物理演算は行わない。図5に示す YCB オブジェクト [5] を対象物体とし、物体認識モデルと姿勢推定モデルに PoseCNN [1] と DenseFusion [2] を用いた。

表1 シミュレーション実験の結果

Object name	Percentage of error less than 2cm[%]		Error reduction[cm]
	start	end	
coffee_can	73.3	96.3	0.28
cracker	93.1	100.0	0.30
sugar	82.6	86.0	0.37
tomato_can	95.6	100.0	0.33
mustard	77.9	99.6	0.55
meat_can	97.1	98.6	0.13
banana	58.7	98.8	0.66
pitcher	58.7	86.9	0.47
bowl	66.5	96.9	0.87
mug	97.9	100.0	0.29

3.1 シミュレーション実験

シミュレーション環境と物体をレンダリングする仮想空間には Gazebo [6] を利用した。簡単化のために、ロボットによる観測は観測位置にカメラのみを生成することで代替し、選択された次の観測位置にカメラを再度生成することでロボットの移動を再現する。また、カメラの初期位置は対象物体から 0.8m の距離で画像中央に対象物体が映る場所とした。図5に示す各物体につき3次元空間の x, y 軸にそれぞれ 60° 、 z 軸に 12° 刻みで回転させた 270 通りの姿勢に対して、提案手法を用いた物体姿勢の推定精度の変化について評価する。

提案手法による視点計画開始時と終了時に推定している物体姿勢 P_{now} とその真値に対する誤差を算出した。開始時と終了時において、姿勢推定誤差が物体を操作する上で許容可能な範囲と考えられる 2cm 未満（文献 [2] 参照）となった割合と、誤差の減少量を表1に示す。提案手法により、すべての物体に対する推定誤差は減少し、推定誤差が把持操作の許容範囲内に含まれる割合は物体平均で 97.6% となった。

3.2 実機実験

実機実験にはトヨタ自動車株式会社の開発する生活支援ロボット Human Support Robot [7] を利用した。対象物体には図5に示す YCB オブジェクトから、cracker, mustard, banana, pitcher, mug の5種類を選択し、物体の形状の都合上、cracker, mustard は2通り、その他は1通りの置き方で設置した。ロボットの初期位置から見える物体面を変化させるために、対象物体を 30° 刻みで回転させた各姿勢に対して提案手法を評価する。また、ロボットの初期位置は対象物体に対して 0.8m 離れた位置とした。

3.2.1 定性的評価

図7に提案手法開始時と終了時の姿勢推定結果を、仮想空間上で可視化した例を示す。仮想空間上のテーブル位置は固定されており、実空間上の真のテーブル位置を反映できていない。そのため、対象物体との相







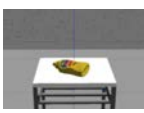
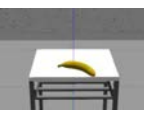

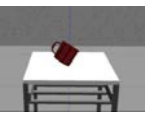

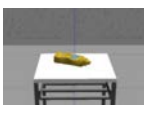

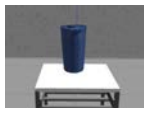
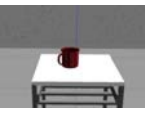
	cracker	mustard	banana	pitcher	mug
true pose					
start pose					
end pose					

図7 実験例:真の姿勢, 開始位置の姿勢推定結果, 終了位置の姿勢推定結果

対位置にずれが生じているが, このテーブルの位置ずれは物体姿勢推定に影響しない。

図より, 視点計画終了時の姿勢推定結果は, 開始時に比べ対象物体の設置姿勢をより正確に反映していることがわかる。物体の表裏が逆転して推定された例も見受けられるが, 物体のテクスチャを考慮せずに物体形状のみから画像の類似度を判定したことが原因であると考えられる。以上より, 提案手法によって推定姿勢精度が改善したことが確認できた。

3.2.2 把持実験による評価

次に, 提案手法によって物体把持精度がどの程度向上するかを検証した。初期位置のみで観測する手法, 直進移動のみによって観測位置を変更する手法, 提案手法の3つの条件において姿勢推定結果を更新し, 最終的に得られた物体姿勢に基づき物体を把持する。把持動作は推定姿勢に対応する掴み方をあらかじめ手動で設計した。図8に各条件における物体把持成功率を示す。物体 mug では把持成功率があまり改善しなかったが, これは物体認識モデルの PoseCNN が mug の取っ手をセグメンテーションできず, 画像の類似度計算において適切な姿勢選択が困難であったことが原因だと考えられる。しかし, 全物体において提案手法により更新された推定姿勢での把持成功率が最も高く, 実空間での把持タスクにおいても提案手法が有用であることが確認された。

4. おわりに

本稿では, 物体姿勢の推定精度を改善するために, 現在位置で推定した物体姿勢に基づいて仮想空間中に物体を配置し, 姿勢推定精度を向上させる観測位置を効率的に選択する視点計画手法を提案した。シミュレーション実験により, 提案手法が物体姿勢の推定精度を改善し, 平均 97.6% の試行で把持操作の許容範囲内で姿勢推定できることを示した。また実ロボットを用いた把持実験により, 実空間においても物体姿勢推定精度が向上することと, 把持タスクで要求される厳密な姿勢推定にも提案手法が有用であることを示した。

今後の展望としてオクルージョンが発生する複雑な環境への適用や複数物体の姿勢推定へ応用を検討する。また仮想空間上での計算効率化により, より高速な視点計画を目指す。

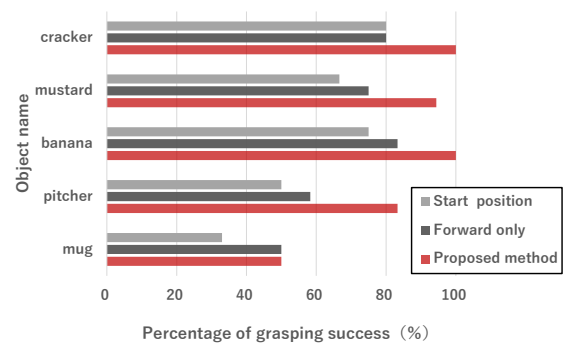


図8 3条件における把持成功率

謝 辞 本研究は JST CREST(JPMJCR15E3) と JST ムーンショット型研究開発事業 (JPMJMS2011) の支援を受けたものです。

参 考 文 献

- [1] Yu Xiang, Tanner Schmidt, Venkatraman Narayanan, and Dieter Fox. PoseCNN: A convolutional neural network for 6D object pose estimation in cluttered scenes. In *Robotics: Science and Systems (RSS)*, 2018.
- [2] Chen Wang, Danfei Xu, Yuke Zhu, Roberto Martín-Martín, Cewu Lu, Li Fei-Fei, and Silvio Savarese. Densefusion: 6d object pose estimation by iterative dense fusion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3343–3352, 2019.
- [3] Christopher McGreavy, Lars Kunze, and Nick Hawes. Next best view planning for object recognition in mobile robotics. *CEUR Workshop Proceedings*, 2017.
- [4] Ying Zhang, Guohui Tian, Jiaying Lu, Mengyang Zhang, and Senyan Zhang. Efficient dynamic object search in home environment by mobile robot: A priori knowledge-based approach. *IEEE Transactions on Vehicular Technology*, Vol. 68, No. 10, pp. 9466–9477, 2019.
- [5] Berk Calli, Arjun Singh, Aaron Walsman, Siddhartha Srinivasa, Pieter Abbeel, and Aaron M Dollar. The ycb object and model set: Towards common benchmarks for manipulation research. In *2015 international conference on advanced robotics (ICAR)*, pp. 510–517. IEEE, 2015.
- [6] Nathan Koenig and Andrew Howard. Design and use paradigms for gazebo, an open-source multi-robot simulator. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566)*, Vol. 3, pp. 2149–2154. IEEE, 2004.
- [7] Takashi Yamamoto, Koji Terada, Akiyoshi Ochiai, Fuminori Saito, Yoshiaki Asahara, and Kazuto Murase. Development of human support robot as the research platform of a domestic mobile manipulator. *ROBOMECH journal*, Vol. 6, No. 1, pp. 1–15, 2019.