

生活支援を想定したロボットマニピュレーションのための物体認識

○橋本学（中京大学），秋月秀一（慶應義塾大学／中京大学），飯塚正樹（中京大学），
鳥居拓耶（中京大学），城亮輔（中京大学），村井俊哉（中京大学）

Object Recognition for Robot Manipulation in the Home Assistance

○Manabu HASHIMOTO †, Shuichi AKIZUKI ‡ †, Masaki IIZUKA †,
Takuya TORII †, Ryosuke TACHI †, Toshiya MURAI †

† 中京大学（Chukyo University） ‡ 慶應義塾大学（Keio University）

Abstract: In this manuscript, we introduce a series of robot vision technologies which can be utilized for home robots. In order to realize practical home robots, recognition of “functions” of everyday objects is required in addition to usual object recognition techniques. Here, for example, “contain” or “grasp” are typical examples of functions. The robot can generate natural and appropriate motion path based on result of such recognition of “functions”, because that will provide beneficial information about usage of objects to be operated.

1. 緒言

ロボットビジョン技術は，生産，物流，家庭等のさまざまな状況下で必要な技術であり，既に普及率の高い分野もあるが，家庭における生活支援ロボットに関しては，まだ発展途上であると言える。

家庭で人間の指示を受けて人間を支援するロボットについては，たとえば「ジュースを持ってきて」のような人間による極めてシンプルな要求を，複雑なロボットマニピュレーションタスクとして実際に実行することが必要である．これを全自動で実行するためには，要求を受けてから実行するまでの手順を生成する必要があるが，生活シーンという日常環境を考えると，ロボットビジョンの観点からは，3次元形状モデル等の個々の対象物に関する事前知識をほとんど使わずにおこなうことが求められるという点において，特段に難度の高い課題となっている。

このような知的ロボットシステムの実現を目指して，我々は，2015年度より，国立研究開発法人新エネルギー・産業技術総合開発機構（NEDO）の支援を受けて，産業技術総合研究所からの再委託事業として，一連の技術開発をおこなっている．本稿では，我々が推進中の物体認識技術開発について，これまでの研究成果のなかの主要なものをまとめて紹介する．具体的には，日用品が持つ「機能」の認識や，適切なロボット動作パラメータ推定のためのロボットビジョン技術について，最新の開発成果を紹介する．以下，まず最初に生活環境におけるロボットビジョンに求められる課題を整理し，これらの具体的な研究成果について説明する。

2. 生活環境におけるロボットのビジョン

2.1 生産・物流・家庭におけるロボットビジョン

図1は，生産，物流，家庭の各場面におけるロボットビジョンの使用例である．製造物が工場で生産され（①），家庭からECにて発注された商品として物流・配送センター（②）にて梱包，輸送され，家庭に届く．そこでは，生活支援ロボット（③）が日用品を把持し，ハンドリングする．このように，いたるところでロボットビジョンが使用される。

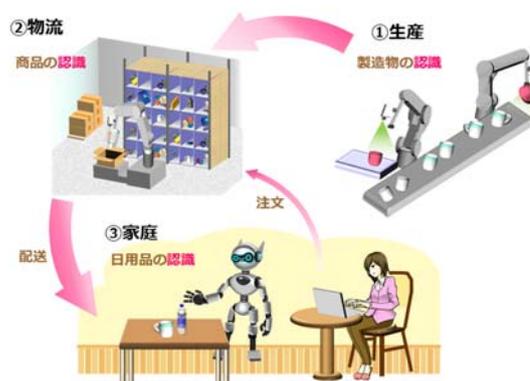


図1. 生産・物流・家庭におけるロボットビジョン

この3つの場面におけるロボットビジョンの種類を図2に示す．生産においては，対象物の形状モデルを使用して物体が「どこにあるか（どのような姿勢か）」を認識する特定物体認識が主タスクになるが，家庭においては形状モデルを使わずに対象物の名称（クラス）を言い当てる一般物体認識が重要である．これらの途中段階である物流・配送センターにおけるビジョンとし

ては、特定物体認識、一般物体認識の両方が必要となってくる。

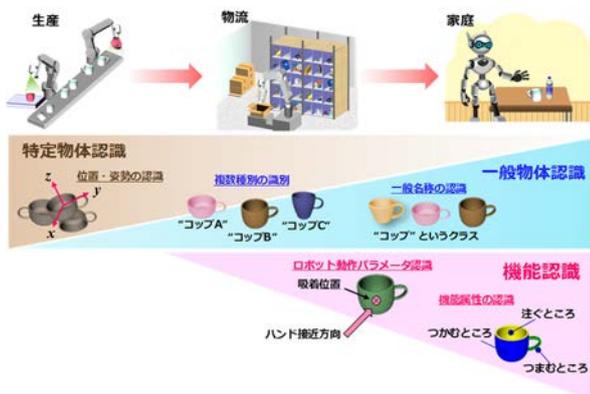


図2. 応用場面ごとのロボットビジョンの種類

2.2 生活支援ロボットの実現に必要な認識機能

生活支援を例にとると、たとえば人間がロボットに対して「コップにジュースを入れて持ってきて」という指示した場合、図3のように、ロボットは、主として、ボトルを持つ、ジュースを注ぐ、コップを持つという3つのサブタスクを実行する。これらはさらに細かく分割できる。たとえば、「ジュースを注ぐ」場合は、コップを見つける、注ぐべき位置を判断する、ボトルを傾けるとい、少なくとも3つの動作を実行する。



図3. 生活支援タスクにおける機能認識の重要性

このうち、コップを見つけるような識別タスクは、いわゆるディープラーニングのような画像ベースの識別器の急激な発展により、かなり容易に実現できるようになってきた。またボトルを傾けるとい基本的ロボット動作生成については、作用点が決まっていれば、さほど難易度は高くない。このことから、これらのサブタスクを実現するためには、「持つべき位置」や「注

ぐべき位置」を判断すること、すなわち、対象物の「機能」を認識するという問題が最も重要であると考えられる。したがって、図2にみられるように、特に家庭内でのロボットシステムにおいては、一般物体認識と同様に、モデルレスを想定した「機能認識」への要求が高くなっていくと考えられる。

3. 対象物の「機能」認識

3.1 日用品の機能

一般に、日用品とは、何らかの特定の機能を有している。その期待されている機能を実現するために、特に局所的な形状が作り込まれていることが少なくない。図4に、日用品が持つ機能の例を示す^[1]。

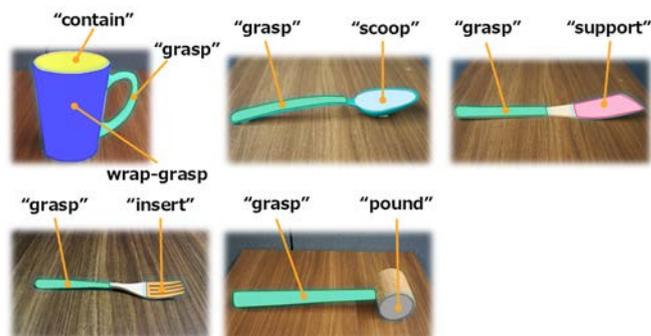


図4. 日用品が持つさまざまな「機能」の例。

例えば、コップには、本来の使用目的である水（液体）を蓄えるということを実現するために、大きくくぼみ形状の部位には”contain”という機能が備わっていると考えられる。同時に、人間が手でつまむことを想定して取って部分には”grasp”という機能が、また胴体部分については人間が手のひらで掴むことができることから、”wrap-grasp”という機能が備わっていると考えることができる。

表1. 機能を決定する主な要因

機能を決定する 主な要因	具体例
形状	いすの support, コップの grasp, コップの contain, ヘラの support, おたまの scoop, スプーンの scoop, 水筒の wrap-grasp, 弁当箱の contain, 洗面器の contain, ハケツの contain, アイロン台の support, シヤワーカーテンの support, ラちむの support, アルミコップの contain, 植木鉢の contain, ベンの grasp, かさの grasp, お盆の support, ゴミ箱の contain, 泡だて器の grasp, 台車の support, ほうきの grasp, かさの wrap-grasp, 洗剤ボトルの wrap-grasp, ドライハの wrap-grasp, 印鑑の grasp, きのりの grasp, メジャーの wrap-grasp, 蛇口の wrap-grasp, フラワースタンドの contain, 脚立の support, 靴べらの support, 灰皿の contain, リモコンラックの contain, はかりの support, カミソリの grasp, ホットプレート の contain, やんばんの contain, コースターの support, 水切りかごの contain
素材	いすの support, はさみの cut
大きさ	いすの support, コップの wrap-grasp
重力	コップの contain, ヘラの support, スプーンの scoop, まな板の support, 橋の contain, 皿の support, 鍋の contain, フライパンの contain, スコップの support, 急須の contain, お盆の support, かごの contain, ボウルの contain, S字フックの hang, 袋の contain, ハケツの contain, ハンガーの hang, 物干し竿の hang, ベーバーホルダの hang, 箸立ての contain, 植木立ての contain, 銅鉄の support, れんげの scoop, しやもじの support, タオル掛けの hang, 洗面器の contain, アイロン台の support, アルミコップの contain, シヤワーホルダの hang, 植木鉢の contain, ゴミ箱の contain, 台車の support, 脚立の support, 灰皿の contain, リモコンラックの contain, はかりの support, ホットプレート の contain, コースターの support, 水切りかごの contain
動き	ナイフの cut, ナイフの sting, フォークの sting, コップの pound, ハンマーの pound, はさみの cut, 箸の pinch, 箸の sting, スプーンの pound, 鍋の pound, フライパンの pound, 歯ブラシの sweep, のごりごの cut, スコップの dig, くしの comb, ピンセットの pinch, クリップの pinch, ビーラーの cut, ハンマーの pound, アイスマシンの sting, 爪楊枝の sting, 寝いすの sting, ペンチの pinch, カッターの cut, 洗面/せみの pinch, ペンの sting, トングの pinch, ベンの sting, かさの sting, ほうきの sting, ほうきの sweep, かさの sting, ドライハの sting, ボタンの pushed, くぎの sting, きのりの sting, カミソリの cut

機能を決定する要因としては、形状以外にも考えられる。例えば、はさみがもつ「cut」は、その実現のためには金属製であることが必要であろう。また、コップがもつ「contain」も、くぼみ状の形状が重力軸と一定の関係を満たさないかぎり（例：上を向いたコップ）、水を蓄えることはできない。このような例を表 1 に示す。

3.2 局所的な形状に基づく機能認識

我々は、3 次元形状に起因して付与されている機能を認識する手法を開発した^{[2]~[5]}。この手法は、図 5 に示されるような機械学習ベースのものである。

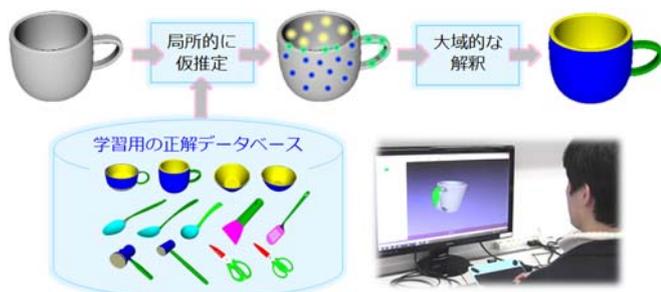


図 5. 機能認識アルゴリズムの概念

まず最初に、学習用の正解データベースを作っておく。3 次元形状モデルに対し、事前に部位ごとに機能ラベルを付与したデータを準備しておく。一方、レンジファインダから得られた未知対象物のポイントクラウドデータに対し、このデータベースを適用して対象物上の局所領域ごとに機能 ID に関する仮ラベルを推定する。具体的には、ランダムフォレストやディープラーニングを利用している。次に、この仮ラベル群を統合し、最終的な機能ラベリングを実現する。

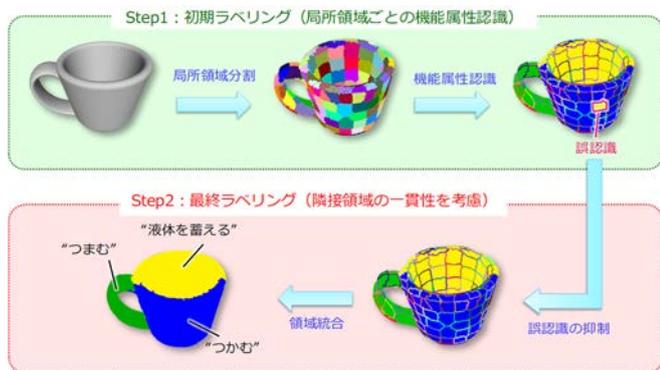


図 6. 仮ラベル統合による機能ラベリングの高信頼化

ここで、初期段階における仮ラベル付与の単位とし

ての局所領域の大きさが重要となる。提案手法では、図 6 に示した程度の大きさを想定している。最終的なラベル領域より十分小さいものの、ポイントクラウドデータを数十点以上を含む程度の領域が望ましい。この局所領域ごとに仮ラベルを付与し、これを統合する過程において、たとえば、隣接領域のラベルは同一であるべきなどの制約を満たすように一部の尤度の低いラベルを修正し、最終的なラベル結果を得る。この方法としては、dense CRF を利用している。

3.3 実験結果

前節で説明した手法による機能認識結果を図 7 に示す。コップやへら（スパチュラ）等に対して、提案手法を適用した結果、STEP1 では局所領域単位の仮ラベルに付与ミスが散見されるが、STEP2 を経てそれらが低減し、真値に近いラベリング結果が得られていることがわかる。

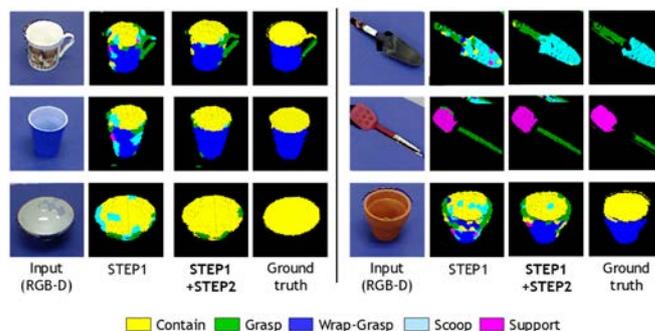


図 7. 機能認識の実験結果

なお、提案手法による処理時間としては、一般的なデスクトップ PC よる CPU 処理により、約 15fps の速度が得られており、これはロボットタスク実行の観点からは、ほぼリアルタイム処理と言ってよい程度の性能である。

4. ロボット動作のためのセンシング

本章では、画像センシング技術を利用して適切なロボット動作パラメータを算出する手法を紹介する。

4.1 プリミティブ形状近似による把持位置決定

把持対象物が未知物体、すなわち対象物に関する形状モデルを利用できない場合でも、ロボットハンドリングのためには、把持位置やハンドのアプローチパラメータを求める必要がある。我々が開発した手法^{[6][7]}は、対象物をいくつかの 3 次元基本図形（プリミティブ形

状)に近似し、事前にプリミティブ形状に対して定義しておいた代表的な把持位置・把持パラメータ(把持位置推定ルール)を未知対象に対して適用するというものである(図8)。

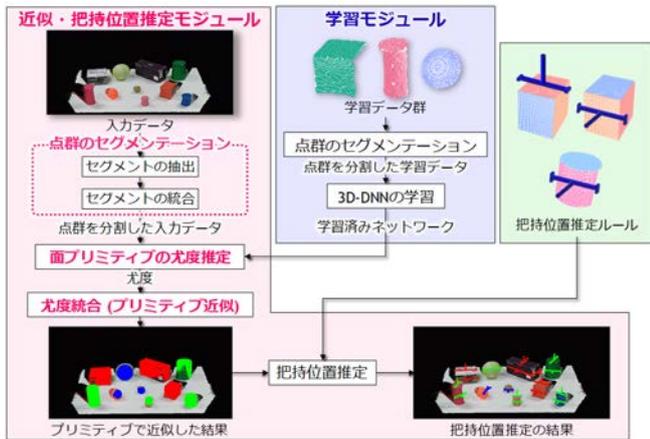


図8. プリミティブ形状近似による把持位置決定

この手法におけるプリミティブ形状近似は、レンジファインダから得られた対象物に関するポイントクラウドデータ(事前にセグメンテーションしておく)に対して面プリミティブを当てはめるという前段処理と、それらを組み合わせて物体プリミティブを当てはめる後段処理の2段階からなる。前段では、3D-DNNを利用して、各面セグメントに対して、平面、円柱面、球面などの面プリミティブラベルの尤度マップを得る。後段では、これらを統合して、直方体、円柱、球の3種の物体プリミティブを当てはめる。なお、日用品の多くが、最終的に本手法によって把持可能であるという意味においてこれら3つのプリミティブで近似可能であることについては、予備調査実験によって確認している。

4.2 機能認識を利用した把持位置推定

前章で紹介した機能認識を、把持パラメータ推定に利用する方法もある。図9に示したように、レンジファインダから得られたRGB-Dデータをもとに、まず機能認識をおこなうと同時に、Dデータを用いた把持パラメータの候補群を得る。この段階では、把持可能な候補が複数得られる。この候補群のなかから最適なパラメータを決定する際に、対象物の局所形状に対して決定済の機能ラベルを利用する。たとえば、"grasp"ラベルが付与された部位に設定された把持位置は正しいが、"contain"ラベルが付与された部位に設定された把持位置は正しくない(好ましくない)と判断される。

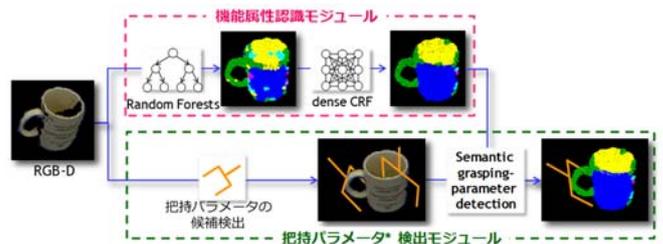


図9. 機能認識を利用した把持位置推定

4.3 リスク最小化を目指した動作パラメータ推定

ロボットピッキングにおいては、把持対象物の周辺に、把持動作を妨げる障害物が存在している場合も少なくない。我々は、このような場面においても有効なロボット動作パラメータ推定手法^{[8][9]}を提案した(図10)。

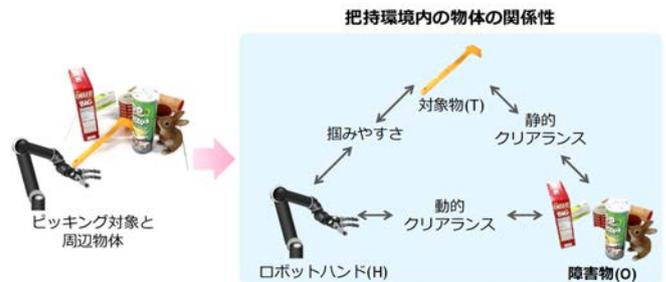


図10. 周辺対象物を考慮した動作パラメータ推定

ここでは、シーン内には、ロボットハンド自身、ピッキング対象物(1個)、それ以外の対象物(すべて障害物とみなす)の3種類しか存在しないと仮定する。したがって、ロボット動作パラメータ推定にあたっては、①ロボットハンドとピッキング対象物との相対的關係(握みやすさ)、②ロボットハンドと障害物との相対的關係(動的クリアランス)、③ピッキング対象物と障害物との相対的關係(静的クリアランス)の3つの関係を指標化し、これらを総合的に最大化することによって、ピッキング動作のリスクを最小化する動作パラメータを求めるという考え方である。

提案手法によって得られた、総合的な把持余裕度マップの例を図11に示す。この把持余裕度をもとに、最も把持余裕が大きい部位に対する動作パラメータを自動決定することにより、ピッキングのリスクが小さい動作が実現する。

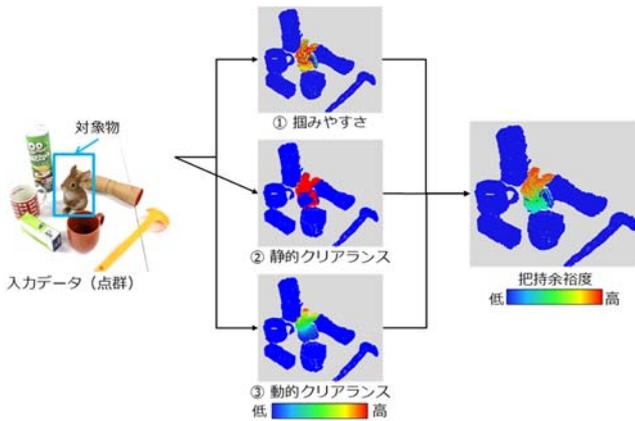


図 11. 提案手法による把持余裕度マップ

5. 結言

本稿では、主として生活支援ロボットへの適用を想定して開発してきた一連の物体認識技術として、対象物の「機能」を認識する技術、および把持位置や把持動作パラメータなどのロボット動作パラメータ推定のためのセンシング応用技術を紹介した。これらはいずれも、3次元形状モデルや、その属性情報などの事前知識が少ない状況下でのロボットマニピュレーションを実現するために必要な技術である。今後は、このようなロボットビジョンとマニピュレーションを円滑につなぐための中間領域の技術がますます必要になると予想される。ひきつづき、これらの性能向上を目指すとともに、シンプルな動作指示をもとに詳細なロボットタスクを自動生成するための技術など、人間とロボットとのインタラクションに関するセンシング情報処理技術を検討していく予定である。

文献

- [1] 山野辺夏樹, 万偉偉, ラミレス・イクシエル, プティ・ダミアン, 辻徳生, 秋月秀一, 橋本学, 永田和之, 原田研介, ロボットマニピュレーション研究におけるアフォーダンスに関するレビュー, 日本ロボット学会誌, Vol. 36, No. 5, pp. 327-337, 2018.
- [2] 飯塚正樹, 秋月秀一, 橋本学, 2.5D データによる日用品の機能属性に基づいたパーツセグメンテーション, 日本ロボット学会学術講演会, 3C1-03, 2017.
- [3] 飯塚正樹, 秋月秀一, 橋本学, denseCRF を用いた日用品に備わる機能属性の高精度認識手法, ビジョン技術の実利用ワークショップ (ViEW2017), IS1-C2, pp. 157-160, 2017.
- [4] Masaki Iizuka, Shuichi Akizuki, Manabu Hashimoto, Affordance-based 3D Feature for Generic Object Recognition, Proceedings of 13th International Conference on Quality Control by Artificial Vision, Vol. 10338, 103380W-1-6, DOI: 10.1117/12.2266917, 2017.
- [5] Masaki Iizuka, Manabu Hashimoto, Detection of Semantic Grasping-Parameter using Part-Affordance Recognition, The

19th International Conference on Research and Education in Mechatronics, pp. 136-140, DOI:10.1109/REM.2018.8421780, 2018.

- [6] Takuya Torii, Manabu Hashimoto, Model-less Estimation Method for Robot Grasping Parameters Using 3D Shape Primitive Approximation, IEEE International Conference on Automation Science and Engineering 2018, pp. 580-585, 2018.
- [7] 鳥居拓耶, 橋本学, ロボットピッキングにおけるモデルレス把持パラメータ推定のための高信頼プリミティブ物体近似, 第18回計測自動制御学会システムインテグレーション部門講演会 (SI2017), 3A4-02, pp. 2144-2148, 2017.
- [8] Kentaro Kozai, Manabu Hashimoto, Determining Robot Grasping-Parameters by Estimating "Picking Risk", International Workshop on Advanced Image Technology, A3-96, 2018.
- [9] 香西健太郎, 橋本学, ピッキング時の空間的余裕を考慮した複雑な環境下における対象物把持パラメータの決定手法, 第18回計測自動制御学会システムインテグレーション部門講演会, 3A4-03, pp. 2149-2154, 2017.

謝辞: この成果の一部は、国立研究開発法人新エネルギー・産業技術総合開発機構 (NEDO) からの委託業務の結果である。