

画像からの質感情報の取得と把持方式の選択

道後千尋¹ 加藤晴久¹ 菅野勝¹

概要：本研究は、ロボットによる物体把持の汎用性を高めることを目標とし、複数方式のマニピュレータを搭載したロボットが、カメラ画像から把持目標の素材を推定し、最も適切な方式のマニピュレータを選択して把持するプロセスを提案する。本稿では素材推定及び適切なマニピュレータの選択を一貫して行うことの出来る機械学習モデルを設計し、その有効性について評価実験を行った。

キーワード：ロボット, 把持動作, 触覚推定, 画像認識, 機械学習

Estimation of Tactile Properties and Suitable Grasping method from Images

CHIHIRO DOGO^{†1} HARUHISA KATO^{†1} MASARU SUGANO^{†1}

1. あらまし

近年、ロボットに生活支援をさせるプロジェクトが多く推進されている[1]. 具体的なタスクは家事や介護、福祉等であるが、調理という作業ひとつをとっても把持すべき対象は多くあり、産業用ロボットとは違う汎用的に作業ができるロボットが必要となる[2]. このような汎用ロボットでは幅広い物体の把持に対応でき、人と同じツールも利用できる多指マニピュレータが搭載されることが多い[3]. しかし多指マニピュレータは同時に複雑な把持位置推定[4]が必要となる。一方で把持位置に依存し難い吸着などの把持方法も存在するが、使用の可否が把持対象の素材に依存するという課題がある。本研究では複数方式のマニピュレータを搭載したロボットが、把持物体の素材を推定し、把持位置依存度の低い手法を優先して選択する把持プロセスの確立を目的とする。可塑性物体や割れ物では接触による転倒や破損の恐れがある為、非接触での素材推定が望ましい。本稿では、画像から質感を定量化し、それを特徴量として把持方式を選択する機械学習モデルを提案し、評価する。

2. 提案手法

本研究の目的は対象の画像から把持方式を決定することである。把持方式の決定には、素材の硬い、柔らかい、滑る、凸凹しているといった質感情報が関係してくる、このことから画像から機械学習によって質感の特徴量を抽出し、その特徴量から把持方式を推定する構成が考えられる。

2.1 先行研究と事前検討

画像から質感の特徴量を得る先行研究として Takahashiらの[5]がある。[5]では入力を画像、出力を触覚センサ値とした Enc-Dec 型の機械学習モデルを使用し、潜在変数として質感の特徴量を得る手法を提案している(表 1 最下段)。しか

表1 従来手法、提案手法、比較手法の概要

	概略図	前半の入出力	後半の入出力
提案手法		画像 ↓ 質感の特徴量	質感の特徴量 ↓ 方式名
比較手法 1		画像 ↓ 質感の特徴量	質感の特徴量 ↓ Yes/No × n個
比較手法 2		画像 ↓ 素材名	素材名 ↓ 方式名
比較手法 3		画像 ↓ 方式名	
従来手法 [5]		画像+センサ値 ↓ 質感の特徴量 ※教師無し学習による 触覚情報の習得のみ	

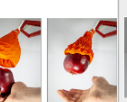
し推定対象は布製品とビニールに限られており、推定する質感も同じ布製品の中で柔らかさを区別する手触りのような繊細な推定を目的としている。そのため本研究の目標である生活支援ロボットでの把持方法決定に適用するには素材のスケールが狭く、粒度が細かい。そこで筆者は[5]で提案されたモデルを用いて事前検討[6]を行い、本研究に適切なモデルとデータセット構成を考察し、表 1 最上段に示す構成を提案した。本稿では考察結果に従い、実際にモデル及びデータセットを作成し、評価を行う。

2.2 提案手法と比較対象となる手法

把持手法を決定する構成には、画像から質感の特徴量を抽出し、その特徴量から把持方式を推定する構成以外にも画像から直接把持方式を推定する等の構成が考えられる。

¹ (株)KDDI 総合研究所
KDDI Research, Inc.

表2 使用できる把持方式の選択肢

名称	磁石	ニードル	吸着(吸盤)	吸着(なじみ)	ジャミング変形	多指(2-5本)
対象	金属	布/粗いもの	滑らかなもの	平面	硬い/厚いもの	制限なし
把持位置依存	小					大
機構	永久磁石又は電磁石で対象をくっつける	針を出し入れし対象を引掛ける	対象との間に真空を作り吸着する	対象を吸引し続け吸着する	対象に密着する形でハンドを変形させ固定する	二本以上の指部を開閉し対象を挟む
製品例	 Pascal	 Fipa	 シュマルツ	 ピアブ・ジャパン	 MIT(2019)	 F-hand NUDO(2018)

本研究では提案手法以外に下記に示す3つの構成を挙げ比較対象とした。各構成の認識手順は表1に示した。

1. 画像から質感を推定し,1方式の可否を推定(比較手法1)
2. 画像から素材名を推定し, 把持方式を推定(比較手法2)
3. 画像から直接,把持方式を推定(比較手法3)

これらの構成のメリット/デメリットを以下に示す。

■比較手法1

各把持方式に対して可否を出力するモデルを,把持方式の数だけ用意する構成。実用化に至っているロボットは[7][8]のように2種の把持方式を搭載するもののみである。従って, 使用可能な把持方式に応じてモデルを組み合わせて構成できるというメリットがある。一方で異種ロボットの連携管理を中央で一元的に行うロボットプラットフォーム[9]のようなシステムでは,作業を命じるロボットを変更することで把持方式の選択肢を増やすことが可能である。この場合は1つのモデルから把持方式に対するスコアを得る事で優先順位を付けられる提案手法が有効となる。

■比較手法2

質感ではなく素材名を推定する構成。比較手法2に対する本手法の利点は学習用データが少量で済むこと,データセットの設計に主観が入りにくいことが挙げられる。具体的には比較手法2で素材名を推定する為には学習用データに素材名をラベリングする必要があるが,素材の選択や粒度は人が決める必要があり,ラベリングを行う人間に十分な知識と規格の統一が求められる。提案手法ではセンサデータを教師データとして用いるため,規格が統一しやすい。また今回のように,最終的な出力(適切な把持方式)が排他的関係ではなく,複数の出力が当てはまる可能性がある問題設定において0/1の教師データを付けるよりも連続値を教師データとすると精度が改善するという研究[10]もある。

■比較手法3

質感の定量化を経由せずに1つの機械学習モデルで推定する構成。比較手法3に対する本手法の利点は異なるロボット/プラットフォームに適用する際に,表1で示すモデル後半部分のみ再学習を行えばよいという運用上の点である。比較手法3は把持方式の選択肢が変更される毎に全体の再学習が必要になる。一方で提案手法に必要なセンサデー

学習パート

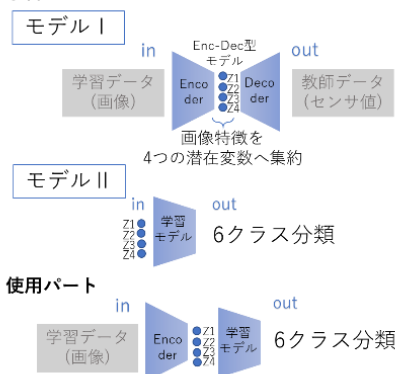


図1 提案方式概要

タを用意する必要はない。

以上のメリット/デメリットから本提案の方針を決定した。各構成に対する比較は4章のモデル評価実験でも行う。

3. 提案モデルの構成

3.1 タスク設定と提案モデル概要

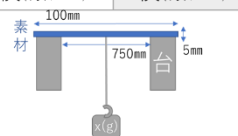
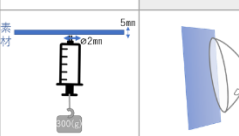
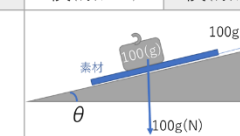
本研究は複数方式のマニピュレータを搭載したハイブリッドハンドロボットでの使用を想定しており,把持に使用できる把持方式の選択肢は表2に示す6種類とする。ロボットは画像を入力として把持方式を選択するが,複数の把持方式が使用可能な場合は把持位置への依存度が小さい方式を優先する。把持対象となる素材は,生活支援ロボットを念頭に日常生活において人が把持する物質の中から図2-1の20種類を選択した。

画像から把持方式を推定する方法は機械学習を使用し,提案モデルは画像から質感の特徴量を算出するモデルIと質感の特徴量から把持方式を推定するモデルIIを独立して学習させる構成(図1)からなる。モデルIでは画像を入力,センサ値を出力としたEnc-Dec型のモデルで潜在変数として質感の特徴量を学習し,モデルIIでは質感の特徴量を入力,把持方式6種を出力とした3層のニューラルネットワークを学習させる。提案モデルはモデルIのエンコーダとモデルIIを結合させ,入力画像から直接把持方式を推定する。

3.2 学習及びテストデータセットの作成

モデルIの学習の為,図2-1に示す20種類の素材の画像(200×200px,モノクロ)とセンサデータを用意した。学習用画像は,撮影距離10cm±3cm以内で各素材10枚の撮影を行い,回転及び拡大縮小によるデータオーギュメンテーションを行い,計1035枚を用意した。センサデータについては,把持方法の決定に寄与すると考えられる力学的な特性と,画像上で物体の判別に寄与すると考えられる表面特性を表す物理量を用意した。具体的には力学特性として強度を表す曲げ弾性,靱性を表す弾性限界,摩擦性に関わる摩擦係数を計測し,表面特性として通気性,吸着性,光沢を計測した。各力学特性のレンジは,人と協働するロボットの出力は人間程抑えられることが多いことから,高齢者の握力(ピンチ力)が5kg±1.5kgである[11]という下限値をとった。

表3 使用したセンサ値と取得方法

名称	曲げ弾力	弾性限界	通気性	吸着性	静止摩擦係	動摩擦係	光沢
方向	縦(特微量1) 横(特微量2)	縦(特微量3) 横(特微量4)	無(特微量5)	無(特微量6)	縦(特微量7) 横(特微量8)	縦(特微量9) 横(特微量10)	縦(特微量11) 横(特微量12)
計測方法							
範囲	0 - 3 kg	0-3kg	0-1000秒	0/1	0-500N	0-500N	0-100GU

素材名 (製品名)	番号	適切方式
スチール (ホチキス針)	1.	磁石
ブリキ(缶)	2.	磁石
布 (タオル)	3.	ニードル
毛糸 (モップ)	4.	ニードル
不織布	5.	ニードル
ポリ (緩衝材)	6.	ニードル
プラ (ファイル)	7.	吸着(吸盤)
紙 (光沢紙)	8.	吸着(吸盤)
樹脂 (床材)	9.	吸着(吸盤)
ガラス	10.	吸着(吸盤)
陶器 (タイル)	11.	吸着(吸盤)
紙 (段ボール)	12.	吸着(なじみ)
木	13.	吸着(なじみ)
プラ (ラップ)	14.	吸着(吸盤)
織物 (畳)	15.	ジャミング
ゴム (消しゴム)	16.	吸着(なじみ)
コルク	17.	ジャミング
紙 (和紙)	18.	多指
フェイク レザー	19.	多指
パルプ (ティッシュ)	20.	多指

図2-1 : 学習に使用した素材

センサデータは表 3 に示す方法で 12 個の値を測定し、それぞれ標準化した。またモデル II の学習の為、それぞれの素材に対して期待する把持方式を図 2-1,2-2 の通りにラベル付けし教師データとした。ラベルは実際に素材を磁石に近づける吸着する等で決定した。テスト用素材には、学習に使用した素材に加えて図 2-2 に示す 15 種の素材の画像を用意した。ロボットのように撮影位置が移動する環境では撮影距離に頑健な認識が求められるため素材 25 と素材 26 は同じ絨毯だが、素材 26 は学習用データより遠い 30cm の距離で撮影している。素材 26 以外のテスト用画像は学習用画像に準じた距離で撮影した。素材 35 はビニールに包装されたマスクであり、透明素材への反応を評価するために用意した。

3.3 モデル詳細と学習環境

モデル I は 4 つの 2D convolutional レイヤーを持つエンコーダと、それぞれ 4 と 32 のノードを持つ全結合層、2 つの 2D convolutional レイヤーを持つデコーダから構成される。またすべてのレイヤーの後に batch normalization を行っている。モデル I 各層のノード数は表 4 に示した通りである。この全結合層のうち 4 ノードを持つ方の出力を潜在変数として、モデル II の入力とした。

表 4 提案モデル I の各レイヤー

	レイヤー	Filters	Kernel_size	Activation
Enc	Conv2D_1	32	(8,8)	Relu
	Conv2D_2	32	(8,8)	Relu
	Conv2D_3	32	(4,4)	Relu
	Conv2D_4	32	(4,4)	Tanh
Dec	Conv2DT	8	(1,2)	Relu
	Conv2DT	1	(2,2)	Tanh

素材名 (製品名)	番号	適切方式
葉	21.	多指
木	22.	吸着(なじみ)
陶器 (カップ)	23.	吸着(吸盤)
布 (キャンバス)	24.	ニードル
布 (絨毯)	25.	ニードル
布 (絨毯)	26.	ニードル
布 (毛布)	27.	ニードル
紙 (水彩紙)	28.	多指
紙 (トレペ)	29.	多指
紙 (ボール紙)	30.	吸着(なじみ)
スポンジ	31.	ニードル
樹脂 (床材)	32.	吸着(吸盤)
結露した 金属	33.	磁石
パルプ (巻紙)	34.	多指
包装された 紙	35.	吸着(吸盤)

図2-2 : 未学習の素材

またモデル II は中間層がノード数 8 の全結合層の 3 層 NN を使用した。学習では optimizer として Adam[12]を使用し、バッチサイズは 15 とした。構成は先行研究[5]をベースとして、センサデータの次元が低い分 Dec のレイヤー数を削減して過学習を防いだ。学習には 64 GB RAM, Intel Core(TM)i7 6700K CPU, GeForce GTX Titan X の PC を使用し、学習時間は両方合わせて 20 分であった。

4. 提案モデルの評価

4.1 提案モデルの精度評価

精度評価は図 2 に示した 35 種の素材を 105 枚で行い、提案モデルによって把持方式を推定した精度は 0.582 であった。正解率の内訳は素材ごとに偏っており、特にニードル方式、吸着(吸盤)方式を割り当てた素材は 8 割の正解率だった。しかし素材 6(緩衝材)、素材 16(消しゴム)ではランダムと同等の精度しか得られなかった。原因としては複層構造の素材 6 と、マットでテクスチャの無い素材 16 ではピントが合わず、特徴が得づらかったものと思われる。一方で殆どテクスチャが無い素材でも反射が大きい素材は認識精度が良い傾向がみられた。図 3-1,3-2 はモデル I の学習で得られた潜在変数 z1~4 を 2 つずつ散布図で描いたものである。同じ素材は平均して描画している。図 3-1 縦軸の負方向の領域に反射の大きい陶器(素材 11,素材 23)やガラス(素材 10)等の、質感でいえば滑らかな素材が集まっており、センサ値から質感を抽出することに成功しているといえる。一方図 3 では距離違いの同じ素材である素材 25 と素材 26 はどの変数でも遠い位置に分布している(図中★)。今回の評価実験では、

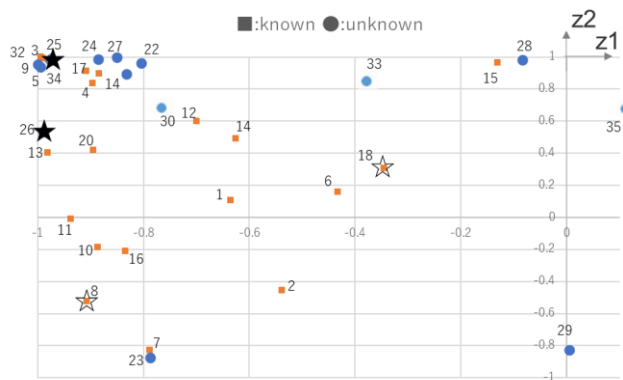


図3-1モデル I の潜在変数z1/2

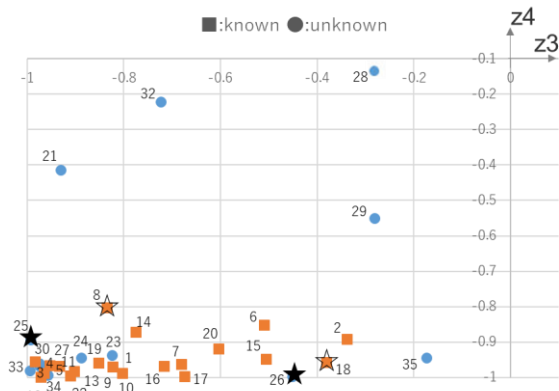


図3-2モデル I の潜在変数z3/4

素材 25 の正解率は 4/5, より遠い距離で撮影した素材 26 は 2/5 と差が出る結果となった. 学習時は拡大縮小を含むデータオーギュメンテーションを行ったものの, 距離変化への対応は不十分であったと考えられる. 素材 35 (包装された紙) は, 正解率は 2/5 であった. 目視で確認したところ, 反射が写った入力画像では正解しており, 写っていない入力画像では紙に近い潜在変数を得ていた. これは画像による把持手法判別の限界と考えられる.

4.2 比較手法 1 との精度比較

比較手法 1 と提案手法は使用環境に応じて使い分ける関係のため, 今回は精度比較を行っていない.

4.3 比較手法 2 との精度比較

本実験の素材では素材 8 (光沢紙) と素材 18 (和紙) は共に紙ではあるが, それぞれ表面が滑らかと凸凹であり, 適切な把持方法が異なり, 前者は吸着できるが, 後者はできないため多指で把持する必要がある. 特に粒度を設計することなく, 遠い値をとっていることが図 3-1, 3-2 両方から読み取れる. テスト用素材では素材 28, 素材 29 が筆記可能な紙だが図 3-2 の横軸の変数が素材 18 と近い値をとっている. 素材 30 は箱等に使用される強度の高い紙であり, 近い値はとっていない. このように同じ紙であっても幅広い性質が存在し, それは把持手法の選択に影響を及ぼすため, 比較手法 2 の学習ラベルの生成には綿密な設計が必要となる. 従ってセンサ値でこれらの素材の違いを自動的に振り分けられる提案手法が優位である.

4.4 比較手法 3 との比較

提案手法は柔軟に構成を変えられる利点があるがセンサデータを用意する必要があり, 単純に優劣は付けられない. 本節では 2 つの手法の精度を比較する. 比較対象として用意したモデルは, モデル I のエンコーダと同じ構成で, 最終層に全結合層を追加し 6 クラス分類モデルとした. 過学習を回避するため全結合層の前に Dropout を挿入している. 各モデルの精度を表 5 に示す. 表より提案手法は学習済みの素材での精度は劣るが, 未知の素材に対する精度が高いことが分かる. 本手法は未知の素材の方がより多く存在する場面で有効と考えられる.

表 5: 提案手法と比較手法 3 の精度比較

	提案手法	直接クラス分類
全体に対する精度	0.582	0.542
学習済み素材に対する精度	0.666	0.7
未学習素材に対する精度	0.471	0.333

5. まとめ

汎用ロボットにおける把持方式の決定について, 画像から質感を定量化した後に方式を選択する機械学習モデルを提案し, 評価した. センサ値を教師データとすることで, 学習用画像に直接ラベルを付ける手法よりも未知の素材への精度が高くなっており, 本構成の利点は確認された. 一方で一部の素材では正答率が高かったものの, ピントぼけの発生しやすいテクスチャの少ない素材や, 撮影距離の離れたテスト画像, 透明な素材などは判断することができず, 全体として精度は低かった. 今後はより適切な学習素材の選別やセンサデータの取得方法を検討し, 精度改善を行いたい.

参考文献

- [1] NEDO 生活支援ロボット実用化プロジェクトの概要, 貞本敦史, 日本ロボット学会誌 Vol. 29 No. 9, pp.760~764(2011)
- [2] 料理ロボットによる食材のマニピュレーション, 工藤俊亮, 日本ロボット学会誌 Vol. 37 No. 6, pp.471~476, 201
- [3] 多指ロボットハンドの制御法, 小沢隆太, 日本ロボット学会誌, vol36 No5, pp306-315(2018)
- [4] Graspability を導入した DCNN による物体把持位置検出, 荒木諒介, 日本ロボット学会誌, vol36 No8, pp559-566(2018)
- [5] Deep Visuo-Tactile Learning: Estimation of Tactile Properties from Images, Kuniyuki Takahashi, ICRA2019, 20-24th May, (2019)
- [6] 把持動作に有用な質感情報を画像から取得する学習モデルの提案, 道後千尋, 情報通信学会ソサエティ大会(2020)
- [7] NimbRo Picking: Versatile Part Handling for Warehouse Automation, Max Schwarz, IEEE ICRA(2017)
- [8] Robotic Pick-and-Place of Novel Objects in Clutter with Multi-Affordance Grasping and Cross-Domain Image Matching, Andy Zeng et al, IEEE ICRA(2018)
- [9] 分散コボット型ロボットシミュレータ・アーキテクチャ, 安藤慶昭ら, 日本ロボット学会誌 Vol. 26 No.5, pp.407~410, (2008)
- [10] 高齢者の手指の触・圧覚, つまみ力, 握力としびれ症状との関係, 赤松公子ら, 日本看護技術学会誌 Vol. 13, No. 2(2013)
- [11] Learning classification models with soft-label information, Nguyen Q, et al. J Am Med Inform Assoc 2014
- [12] "Adam: A Method For Stochastic Optimization, D. Kingma., (2015)