Vol.2018-CVIM-211 No.11
2018/3/2

物体のパーツ形状と持ち方の共起性に基づく把持パタンの推定

川上 拓也[1,a)]　松尾 直志[1]　島田 伸敬[1]

概要：物体の機能とそれを把持する人間の手の姿勢や形状には密接な関わりがある．本研究では，教師なしの Shift Invariant Sparse Auto-encoder により創発的に獲得した持ち方の記述パラメータ空間（持ち方パラメータ空間）を用いて物体の持ち方を定量的に記述する．さらに，物体は特有の持ち方を持つ形状パーツの組み合わせであると仮定し，パーツと持ち方の関係性（把持パタン）を学習させる．学習によって獲得したモデルを用いた実験では，持ち方パラメータ空間が把持パタンの類似性や相違度を定量的に記述することができるかを確認した．また，未知の物体のみを入力としてその部分ごとの把持パタンを推定した．さらに，ロボットによる物体把持を視野に入れ，部分ごとの想起手形状を統合することによって，持ち方ごとの手の全体像の想起を行った．

Grasping Pattern Estimation Based on Co-occurrence of Object and Hand Shape

Takuya Kawakami[1,a)]　Tadashi Matsuo[1]　Nobutaka Shimada[1]

1. はじめに

1.1 研究背景

外観ベースの一般物体認識の分野では，物体は日々新しい形状のものが生成されるため，物体パーツと把持手形状の組み合わせをロボットに登録しておくのは不可能であるため，難しいとされている．対照的に，物体の機能は人間が扱う多くの物体で共通しており，そのバリエーションは比較的少ない．したがって，物体の機能ベースの一般物体認識は有用だと考えられる．物体の機能ベースの一般物体認識の有用性は既に示されているが [1][2]，各物体カテゴリに対して物体の機能を手動で定義している．多数の物体に対する認識を行う際には，物体にに機能ラベルを手動で割り当てることなく，自動的に抽出できることが望ましい．

人間が把持を行う物体は様々な機能を持っている．人間は物体を把持する際に，その物体の機能に応じて手の形を変えて把持を行う．例えばスプレーなら引き金近くにある細くなっている部分を人差し指以外の四本の指で握り，引き金部分に人差し指を掛けるような持ち方をする．これは，スプレーが引き金を引くことにより液体を噴射するという機能を発現しやすくするためにデザインされた形状であるためである．同様の機能を持った物体であれば同様の形状のパーツが備わっており，人間はそのパーツを認識するとその形状に適した持ち方で物体を把持する．このように物体の機能がその物体と手のインタラクションに密接に関連している [3]．この物体と手のインタラクションの種類は文献 [4]，[5] で正確に分析されている．したがって，外観ベースの認識において，機能ベースの分類のためには手と物体のインタラクションが有用だと考えられる．この” 物体パーツ” と” 人間がそれを把持する際の持ち方’ の共起性を利用して物体のパーツから持ち方を想起し，その持ち方を真似してロボットが物体把持を行うことができれば，人間の様に物体を認識した際に適切な把持が行えるロボットを作ることができる．

[1] 立命館大学大学院 情報理工学研究科
[a)] tkawakami@i.ci.ritsumei.ac.jp

1.2 提案手法

一般物体認識の分野の課題の一つとして学習用のデータセットを収集するのが困難という点がある．それを解消するために，人間の日常生活を観察することによって学習用のデータセットを収集する仕組みが必要である．このデータセットというのは人間が把持を行うパーツ形状とそのパーツの把持手形状の画像である．しかし，データセットを収集するシーンを人間がテーブルに座って物体を把持しているシーンに限定したとしても，どの位置に物体があって，どのタイミングで物体にアプローチするかという事を認識しなければならない．室内シーンイベントを監視・記録するシステムに，室内ロギングシステム [6] というものがある．このモニタリングシステムは Kinect v2 で撮影したシーンからイベントを階層的に検知して，そのイベントおよびシーン画像をロギングするシステムである．これを利用すれば人間が物体にアプローチしたタイミングでその物体を矩形でで切り抜いた深度画像を取得することができる．

本稿では，低次元の空間に手と物体のインタラクション（持ち方）を表す” 持ち方パラメータ” というベクトルを記述するシステムを提案する．持ち方パラメータはあらかじめ定義された物体の持ち方を連続的に記述したものである．提案手法は，教師なし学習によって持ち方パラメータを記述することが可能である．物体の機能が持ち方と密接な関わりがあると仮定すると，持ち方パラメータは物体機能のモデリングとして有用であると考える．持ち方パラメータの数値表現については，持ち方パラメータ空間というものを提案する．この空間は教師なしの特徴抽出法である畳み込みオートエンコーダ (CAE)[7] によって作成される．CAE の学習を行う際に，パラメータ空間内で類似の持ち方をクラスタリングするため，スパース項を導入する．CAE の学習はマグカップ，コップ（取っ手なし），スプレーなどの典型的な機能を持つ物体に対する手と物体のインタラクションの外観に基づいている．従って学習すべき画像は，潜在的な属性である外観そのものと手と物体のセグメンテーション画像を含む” 把持画像” である．その後，畳み込みニューラルネット (CNN) を用いて，物体の外観と持ち方パラメータの共起性（把持パタン）を想起するモデル（把持パタン想起モデル）を作成する．このようにして，物体が写った画像のみから持ち方を想起することができる．

2. 物体画像を用いた把持パタン想起モデル

本稿では物体画像から物体の持ち方を表す記述子である持ち方パラメータを想起するモデルを提案する．本節では，その持ち方パラメータとそれを想起するモデルについて説明する．

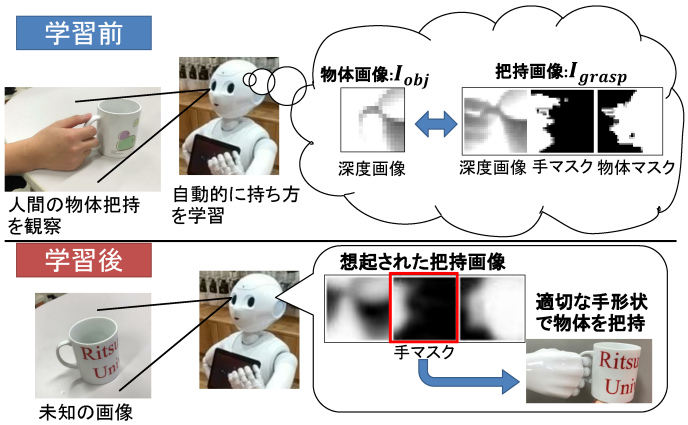


図 1: 人間の物体把持シーンから持ち方を学習するロボット

2.1 持ち方パラメータ

物体の機能はその物体の持ち方と密接な関わりがあるため，物体の持ち方は機能を記述するために有用である．物体の形状と把持する際の手形状は連続的な変化を持っている．そのため，持ち方パラメータはこのような変化を連続的に反映すべきである．本稿では，物体把持の外観を符号化することによって，” 持ち方パラメータ” ベクトルを生成する．

ここで問題になるのは，外観から持ち方パラメータへのマッピング方法である．マッピングは次の条件を満たす必要がある．

A). マッピングは，インタラクションの重要な情報のみを抽出すること． インタラクションに関係しない物体の形状は無視すべきである．

B). マッピングは，事前に手動で対話を分類することなく，ラベルなしの外観セットからの学習が可能であること．

C). 異なるインタラクションに対応する画像は，空間上で離れたパラメータとしてマッピングされること． 2 つの相互作用の差異は，対応するパラメータ間の数値距離に反映されるべきである．

D). 類似のインタラクションに対応する画像は，物体のサイズまたは形状が異なり画像内で互いにわずかにずれる場合でも密接な位置にパラメータがマッピングされること．

E). エッジやグリップなどの空間的に局所的な特徴は，複数の相互作用に共通し相互作用を区別するのに有効である．そのような有用な特徴はを外観から自動的に抽出すること

把持画像の持つ本質的な情報は，エンコーダとデコーダを用いたオートエンコーダ法 [8][9] によって抽出することができる．エンコーダは，入力をより低い次元のコードに変換し，デコーダはコードから基の入力をほぼ復元する．両方の要素は複数の入力に関して可能な限り正確に復元されるよう学習される．この制約の下で，エンコーダは入力復元に必要な主成分の数値表現を生成する．さらに，エンコーダ及びデコータは，教師ラベルなしのベクトル (上記

条件 B を満たす) で学習を行うことができる．持ち方パラメータから外観を復元できる場合，パラメータには持ち方の情報が含まれる．以上により，条件 A と条件 B を満たすマッピングはオートエンコーダ法によって達成される．

条件 C を満たす為に，特定の持ち方に対応するパラメータを空間上で密集させ，他の持ち方に対応するパラメータをその集合から分離する．持ち方を指定するラベルを各外観情報に付けることができる場合，2 つの異なる持ち方パラメータが遠くに配置されるようにマッピングに制約をかけることができる．しかし，条件 B を満たすために，持ち方を指定する重要な要素はラベル無しで抽出しなければならない．これまでの研究で，ラベルの無いベクトル集合の中の重要な要素は，スパース符号化法 [10]〜[14] によって発見されている．しかし，これらの方法は，追加の不等式，または等価式の制約を必要とする．この問題を解決するために，オートエンコーダにスパース制約を導入する．スパースオートエンコーダは以前から提案されているが [9]，[15] の方法はモデルを訓練する際に不等価制約を必要とする．

そこで，オートエンコーダに等価，または不等価式の制約を必要としない，スパース項を導入する．これは，非線形活性化関数を有するニューラルネットワークの層が含まれた一般的なＣＮＮベースのオートエンコーダに適用可能である．ＣＮＮは，画像から空間的に局所的な特徴を一様に抽出する畳み込みフィルタからなるので，抽出された局所的特徴は位置に依存しない (条件 D，条件 E を満たす)．さらに，ＣＮＮフィルタは，教師なし (条件 B を満たす) で学習することができる．

2.2 把持パタン想起モデル作成手順

本稿では，機械（ロボット）が人間の物体把持の様子を観察することによって自動的に物体形状と持ち方（把持画像）の関係を学習し，未知の物体に対しても学習時の経験を基に適切な持ち方を想起する事を目的としている．手順としてはまず，図 2 に示す通り，Auto-encoder を用いて把持画像から持ち方パラメータが写像される空間の学習を行う．このモデルは I_{grasp} を 30 次元の Descriptor にする Encoder 部分と Descriptor から想起画像 $D(E(I_{grasp}))$ を復元する Decoder 部分に分けられる．

次に図 3 に示す通り，Auto-encoder の学習結果である持ち方パラメータを教師とし，CNN を用いて物体のみ画像と持ち方パラメータの関係を学習させる．まず I_{grasp} を Encoder に入力し作成した Descriptor を $P_{teacher}$ とする．これと CNN モデルの出力 P_{recall} の平均二乗誤差を最小にするように回帰学習を行う．これにより，学習に使用していない物体でも学習済みの物体の持ち方の知識から想起することができる．その後，図 4 に示す通り，作成した学習済みの CNN モデルを用いて学習に使用していない物体の持ち方パラメータ を作成し，Decoder を使ってその物体の把持画像を想起する．物体全体ではなくパーツに注目するため，物体画像，把持画像共に 64px×64px から 32px ×32px のパッチ画像を切り出して学習に用いた．

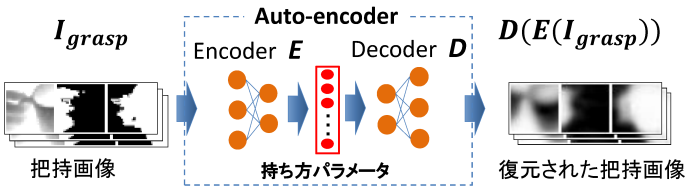


図 2: Auto-encoder を用いた把持画像の復元学習概要図

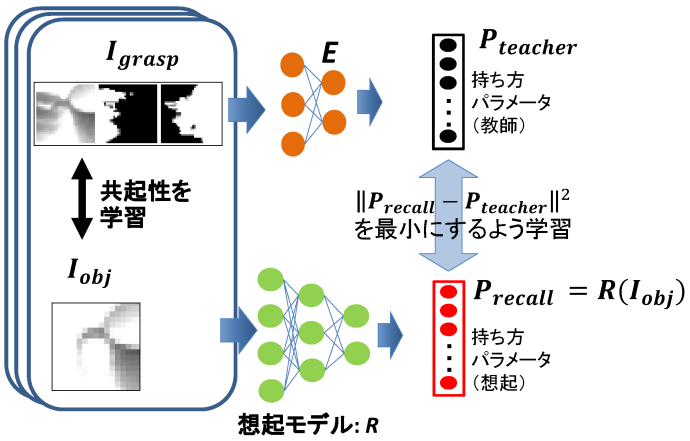


図 3: CNN 想起モデルを用いた物体と持ち方パラメータ学習概要図

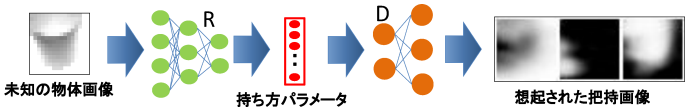


図 4: 未知物体から把持画像を想起する手順の概要図

3. 人の物体把持シーンの観察に基づく学習用把持画像収集

学習に使用する把持画像 (深度画像, 手のマスク画像，物体のマスク画像の 3 チャンネルから成る画像) 作成の手順を図 5 に示す．本研究では，人間の日常生活を観察することによって学習データの収集を行う．撮影には深度画像の取得が行える kinect v2 を用いる．撮影するシーンは，人間が物体に触れていない画像を初期フレームとし，そこから物体を把持し，持ち上げるというシーンとする．このシーンを点群として撮影し，3 次元的なトリミングと平面除去を行い，手と物体以外の点を削除する．次に ICP アルゴリズムを用いて，初期フレームの物体点群を目標として入力点群の位置合わせを行う．その後，Nearest Neighbor 法を用いて，位置合わせ後の入力点群から初期フレームの物体点群に近い点を物体点群に，それ以外を手の点群とす

る．この物体と手の点群を用いて把持画像，手のマスク画像，物体のマスク画像を作成し，それを 3 チャンネルにまとめたものを把持画像とする．同時に初期フレームの深度画像を物体画像として保存しておく．この 3 チャンネルからなる把持画像と物体画像の組み合わせを一対とし，Auto-encoder，CNN を用いた学習で使用する．

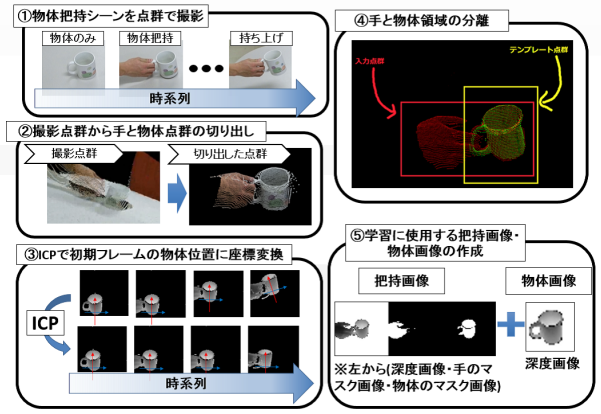


図 5: 把持画像作成手順画像

4. 把持パタン想起実験

本章では物体画像から把持パタン（持ち方）を想起する実験に関して説明する．実験に使用した物体は，図 6 に示す通りマグカップ，(取っ手無し) コップ，ボール，スプレーの 4 カテゴリ，4 種類の計 16 物体である．それぞれの物体に対して，把持画像を約 100 枚ずつ作成し，把持パタンの想起モデルの学習を行った．

図 6: 学習に使用した 16 物体

4.1 Auto-encoder のモデル構造

まず，把持画像から持ち方パラメータを作成するための Auto-encoder について説明する．本実験で使用した Auto-encoder のモデル構造を下記に示す．

- Encoder
  - 畳み込み層 (フィルタサイズ : 9 × 9)[32 × 32 × 3 → 24 × 24 × 16]
  - Tanh
  - L2 プーリング [24 × 24 × 16 → 12 × 12 × 16]
  - Tanh
  - Reshape[12 × 12 × 16 → 2304]
  - 線形結合 [2304 → 1500]
  - Tanh
  - 線形結合 [1500 → 150]
  - Tanh
  - 線形結合 [150 → 30]
  - Tanh
- Decoder
  - 線形結合 [30 → 150]
  - Tanh
  - 線形結合 [150 → 1500]
  - Tanh
  - 線形結合 [1500 → 3072]
  - Tanh

この Auto-encoder の入力が画像であるため，線形結合の前に畳み込み層とプーリング層を挟んでいる．畳み込み層は画像内の特徴を抽出する効果を，プーリング層はその特徴の細かい位置ずれを抑制する効果を狙ってモデルに組み込んでいる．また，本実験では入力画像の細かい位置ずれによる復元画像のぼやけを抑制する為に Shift Invariant Sparse Auto-encoder を使用する．[16] この Auto-encoder は入力を少しシフトさせた画像群からすべて同じ復元画像を出力するような誤差関数を使って学習を行う．物体位置は ICP による位置合わせを行っているが，手形状に関してはほぼ同じ持ち方でも画像中では細かい位置ずれが生じるため，この Auto-encoder を採用した．

4.2 微小領域ごとの把持画像の学習

本実験では物体画像と把持画像の一部を切り取って学習を行う．意図としては，物体はパーツ毎に持ち方が異なっており物体全体に対して一つの持ち方を学習するのではなく，物体のパーツ毎に持ち方を学習する必要がある．微小領域ごとに学習することにより，物体全体ではなくパーツに焦点を当てて学習できるため，今回は物体の一部を切り取ったパッチ画像を入力として学習を行った．図 7 にあるように，物体画像と把持画像の物体位置は同じなので，両画像の同じ場所を切り取り，それぞれを入力として Auto-encoder と CNN の学習を行っている．

4.3 Auto-encoder による持ち方パラメータ空間の学習結果

まず，Auto-encoder による把持画像の復元結果を確認する．図 8 に各物体の復元画像を示す．Shift Invariant Sparse Auto-encoder を採用したおかげで，全体的に鮮明

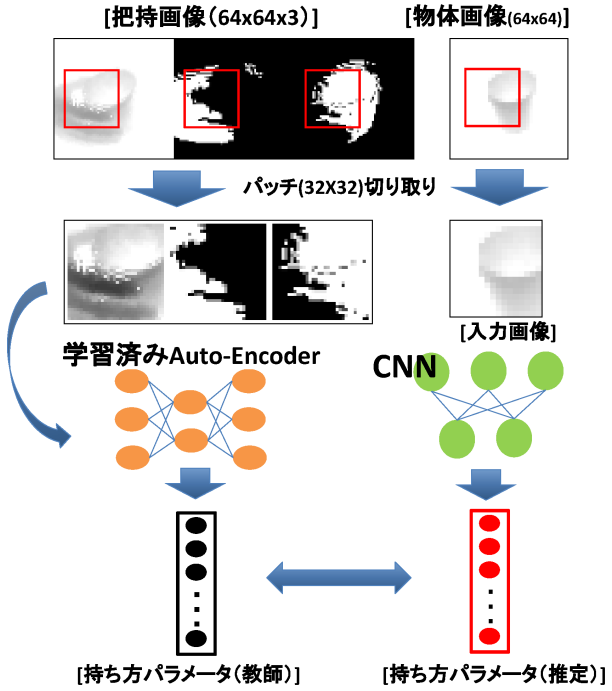


図 7: 局所領域の切り取りについて

に想起されており，特に手のマスク画像については指の一本一本まではっきりと認識できるほど想起されていることが分かる．

次に，Shift Invariant Sparse Auto-encoder の学習により獲得した持ち方パラメータが物体のカテゴリごとに空間上でどの程度分かれているかを確認する．図 9 に持ち方パラメータ空間の第一主成分，第二主成分の分布を示した．同じ物体は持ち方パラメータ空間上でも概ね近い位置にプロットされていることが分かる．また別物体であっても，同じカテゴリであれば第一，第二主成分空間上で近い位置に集まっている．

この学習済み Auto-encoder で抽出した持ち方パラメータを教師として次節の物体画像から持ち方を想起する実験を行う．

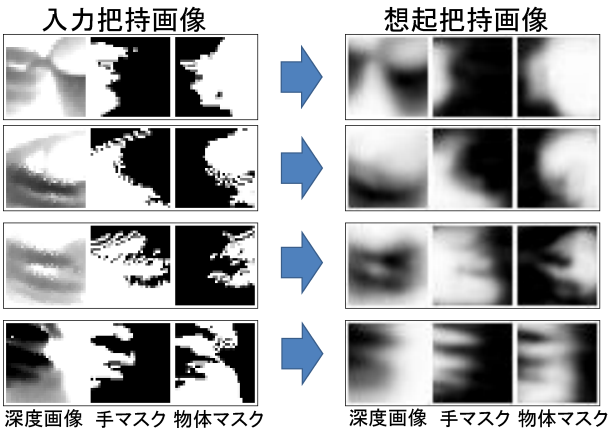


図 8: Shift invariant sparse auto-encoder から復元された把持画像

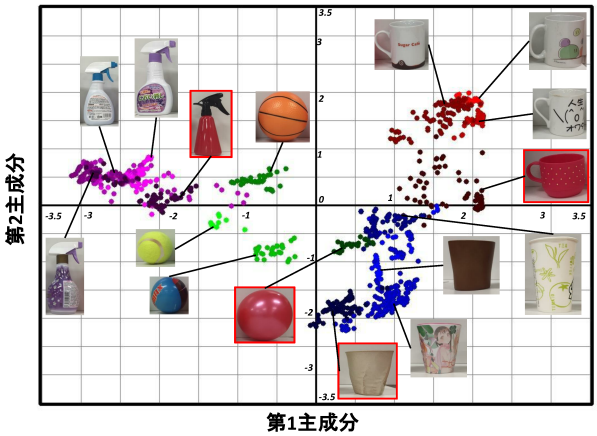


図 9: 把持画像から抽出した持ち方パラメータの分布

4.4 CNN 想起モデルの構造

次に，物体画像のパッチ画像から持ち方パラメータを想起させる CNN について説明する．実験に使用した CNN のモデル構造を下記に示す．

- 畳み込み層 (フィルタサイズ : 5 × 5)[32 × 32 × 3 → 28 × 28 × 16]
- Tanh
- L2 プーリング [28 × 28 × 16 → 14 × 14 × 16]
- 減算正規化
- 畳み込み層 (フィルタサイズ : 5 × 5)[14 × 14 × 16 → 10 × 10 × 64]
- Tanh
- L2 プーリング [10 × 10 × 64 → 5 × 5 × 64]
- 減算正規化
- Reshape[5 × 5 × 64 → 1600]
- 線形結合 [1600 → 1500]
- 線形結合 [1500 → 30]

この CNN では 64px × 64px × 3 チャンネルの画像から 32px × 32px × 3 チャンネルの画像をパッチ画像として切り取って入力としている．最終的には 30 次元の持ち方パラメータを想起するため，出力層は 30 ノードとしている．

4.5 CNN 想起モデルによる把持画像の想起結果

図 10 にトレーニング画像から想起された持ち方パラメータから Decoder を用いて作成した復元画像を示す．
マグカップに関しては，深度画像のコップの淵の位置は異なるものの，手マスクの形状や物体マスクの取っ手部分はしっかり想起されている．コップ，ボールに関しては物体マスクに Auto-encoder による復元の際には見られなかったマグカップの取っ手部分のようなでっぱりが出ていた．深度画像もややマグカップのものと似ているが，手マスクは適切な形状で想起された．スプレーに関しては深度画像，手マスク，物体マスク共に適切な形状で想起された．

次に，図 11 にテスト画像から想起された持ち方パラメー

タから Decoder を用いて作成した復元画像を示す．
マグカップは，淵の部分の深度画像が物体画像の淵の位置と異なって想起されているが，手形状は取っ手の部分を把持しているような手マスク画像が想起された．コップは胴の部分を包むような手形状が想起され，ボールは全体を覆うような手形状が想起されている為，概ね良い結果であった．スプレーに関しては指の形が認識できる程度には良く想起されていたが，物体マスクが入力の物体形状と大きく異なっていた．

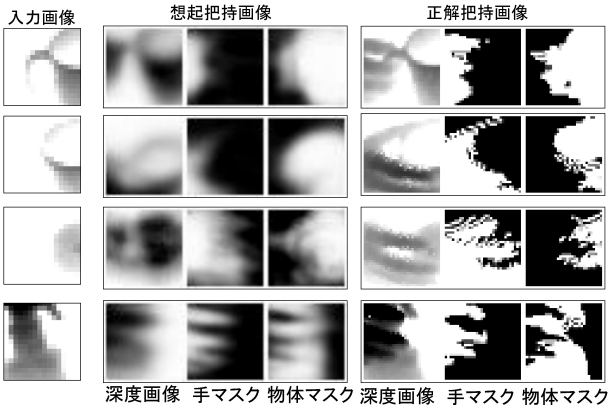


図 10: 物体画像から想起された把持画像 (train)

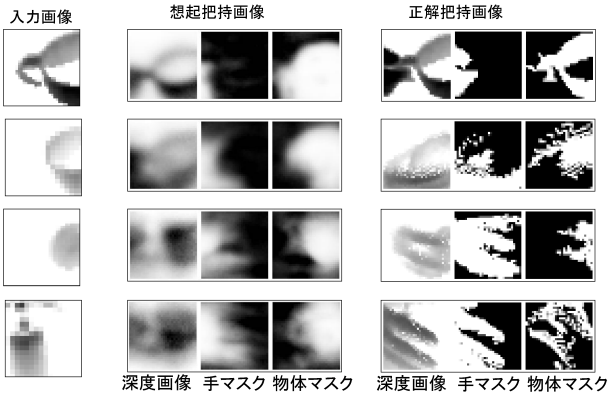


図 11: 物体画像から想起された把持画像 (test)

4.6 重要パーツの有無による持ち方の変化

同物体であってもパーツ毎に持ち方が変化するかを確認するため，一つの物体画像から異なる二カ所をパッチ画像として切り出し，想起を行った．図 12 にその結果を示す．今回はマグカップの「取っ手を含めたパッチ画像」と「取っ手を隠したパッチ画像」からそれぞれ把持画像の想起を行った．「取っ手のあるパッチ画像」に関しては取っ手を握るような手形状が想起され，「取っ手を隠したパッチ画像」に関しては胴の下部を包むような持ち方が想起された．また，マグカップの取っ手を隠したパッチ画像とコップの似たような部分のパッチ画像を比較したところ，概ね同じような手形状が想起された．この結果から，学習モデルがマグカップの重要なパーツである取っ手を認識しそのパーツに適切な手形状を想起していること，マグカップの胴の部分のような学習していないパーツに関しても他物体の持ち方を基に適切な持ち方を想起していることが確認できた．

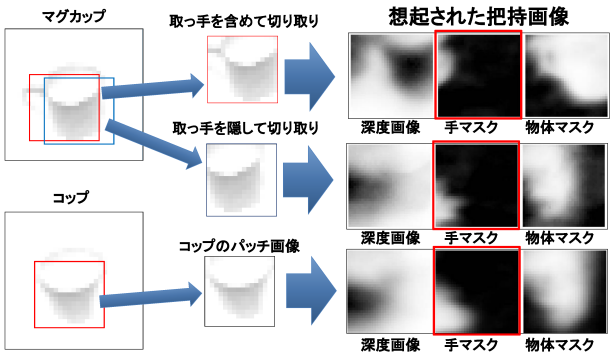


図 12: 重要なパーツの有無による想起把持画像の違い

5. 複数の持ち方を有する物体のパーツ毎の把持手形状想起実験

5.1 パッチ画像の張り合わせによる手の全体領域の想起

現在の想起モデルは物体画像の微小領域に対する持ち方を想起するものである．したがって，手の全体像を想起するためにはその物体画像の全てのパッチ画像から想起した把持画像を張り合わせて一枚の画像にする必要がある．想起されたパッチ把持画像のうち手マスク画像を張り合わせ，手の確率マップとし，それを深度画像の張り合わせ画像にかけることにより，手の全体像の画像を作成した．

5.2 学習に使用した物体の持ち方について

図 13 に学習に使用した持ち方の一覧を示す．学習に使用した物体カテゴリは 4 章で使用したマグカップ，コップ（取っ手無し），ボール，スプレーの 4 種類である．持ち方はそれぞれのカテゴリにつき 2 種類ずつ行った．マグカップの持ち方は取っ手を掴む持ち方，淵を掴む持ち方の 2 種類である．コップの持ち方は筒の部分を掴む持ち方，淵を掴む持ち方の 2 種類である．ボールの持ち方は横から全体を包むように掴む持ち方，上から覆うように掴む持ち方の 2 種類である．スプレーの持ち方は引き金に人差し指をかけて他の指で首の部分を掴む持ち方，下部の膨らんだ部分を掴む持ち方の 2 種類である．

5.3 持ち方のクラスタリングに基づく手の全体領域の想起

1 つの物体に対して複数のパーツに対する持ち方を学習していた場合，パーツ毎の持ち方が混ざった状態で把持画像が想起されてしまう．そこで，図 14 のように，全パッチの持ち方パラメータを空間上でクラスタリングし，同様の持ち方をするパッチ画像の手形状のみを想起し張り合わせる必要がある．今回はマグカップの取っ手を握るような持

図 13: 学習に使用した持ち方一覧

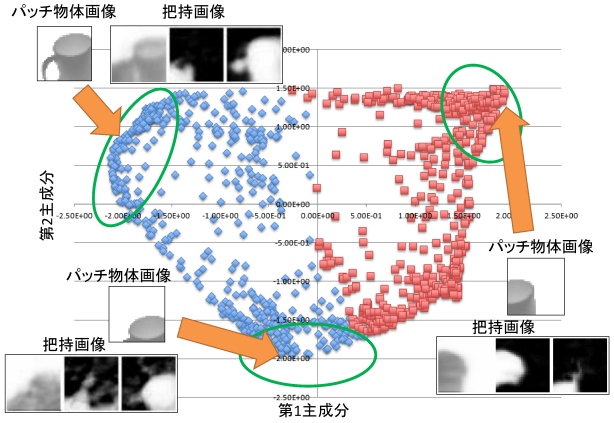
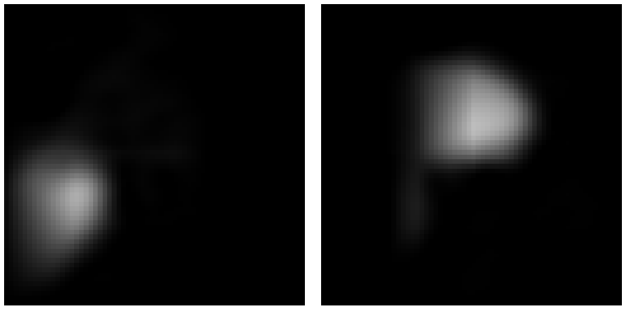


図 14: k-means でクラス分けされた持ち方パラメータの分布



(a) マグカップ持ち方 1　　(b) マグカップ持ち方 2

図 15: 手領域の確率マップ

ち方，淵を掴むような持ち方の二つの持ち方を学習させた想起モデルを用いてパーツ毎の把持手形状の想起を行った．

事前実験として，張り合わせを行う前に $64px \times 64px$ のマグカップ画像から切り取り可能なパッチ画像の持ち方パラメータの分布を確認した．図 14 にマグカップの全微小領域から作成した持ち方パラメータをクラスタリングした分布を示す．クラスタリングには $k = 2$ の k-means 法を用いた．分布を見ると，持ち方パラメータが大きく分けて 3 つの集団に分かれていることが確認できる．これは，取っ手を掴む持ち方（クラス 1），淵を掴む持ち方（クラス 2），そのどちらにも属さない持ち方（クラス 3）の 3 つである．k-means 法は初期値をランダムで設定するため，クラス 3 付近が初期値になると，学習させた持ち方ごとのクラスタリングが行えないため，今回は学習させた持ち方パラメータ付近を初期値としてクラスタリングを行った．

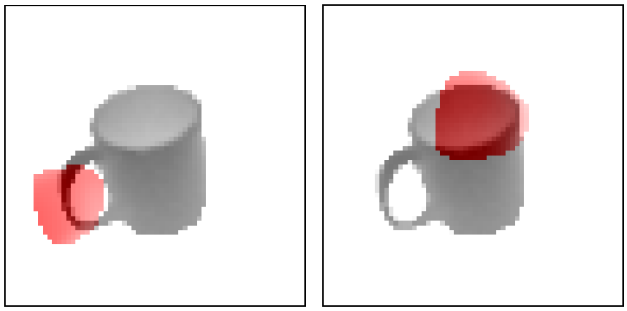
5.4 パッチ画像張り合わせによる手の全体の想起実験

微小領域の想起画像の張り合わせに関しては，まず想起された把持画像の手マスクを持ち方クラスごとに張り合わせ，手マスクの確率マップのようなものを作成した．その確率マップと深度画像を張り合わせた画像の積を手の全体像とした．図 15 に持ち方ごとの手マスクの確率マップを示す．また，図 16 に想起した手の全体画像を示す．

想起された手形状を見るとクラス 1 は取っ手を握るような形状，クラス 2 は淵を掴むような形状が想起された．このように k の数と初期値の場所を指定する必要はあったが，パーツ毎の把持手形状を想起する上で持ち方パラメータのクラスタリングは有用である．クラス 3 のようなクラスタができた原因としては，クラスタ 3 は二つのパーツをどちらも一部含んだパッチ画像から想起されているため，各パーツに対する把持画像の中間的な画像を想起した事が考えられる．今後はパッチ画像に対して複数の持ち方が考えられる際に中間的な持ち方を想起するのではなく，複数の持ち方を想起するモデルを考案する必要がある．

6. おわりに

本稿では，物体の機能とそれを把持する人間の手の姿勢や形状には密接な関わりがあるという特性をもとに，物体把持時の手形状から物体機能の記述，またその記述子（持ち方パラメータ）を用いて物体から適切な手の把持形状の想起を行う手法を提案した．物体把持時の深度画像と手，物体のマスク画像（把持画像）から教師なしの Shift Invariant Auto-encoder により創発的に獲得した持ち方パラメータによって物体のパーツごとの持ち方を定量的に記述した．



(a) マグカップ持ち方 1　(b) マグカップ持ち方 2

図 16: 手の全体像の想起

また，物体画像から対応する把持形状を想起するモデルの作成を行い，物体のパーツに対応した手形状の想起を行った．実験により，提案した想起モデルは学習に用いていないパーツを入力しても似たような形状のパーツを学習していればそれに対応した手形状を獲得することができる事が確認できた．更に，複数の持ち方を有する物体の微小領域ごとの持ち方パラメータをクラスタリングし，パーツ毎に手形状を想起する実験を行った．実験により，複数パーツを含む微小領域では各パーツの持ち方の中間的な持ち方が想起された．今後は，一つの入力に対し複数の回答を想起するモデルを考案し，ロボットによる物体把持を目指す．

参考文献

[1] L. Stark and K. Bowyer, "Achieving generalized object recognition through reasoning about association of function to structure," IEEE Trans. Pattern Anal. Mach. Intell., vol.13, no.10, pp.1097-1104, Oct. 1991.
[2] K. Bowyer, M. Sutton, and L. Stark, "Object recognition through reasoning about functionality: A survey of related work," Object Categorization: Computer and Human Vision Perspectives, p.129, 2009.
[3] D. Bub and M. Masson, "Gestural knowledge evoked by objects as part of conceptual representations," Aphasiology, vol.20, no.9, pp.1112-1124, 2006.
[4] J.R. Napier, "The prehensile movements of the human hand," J. Bone and Joint Surgery, vol.38, no.4, pp.902-913, 1956.
[5] N. Kamakura, *Shape of hand and Hand motion*. Ishiyaku Publishers, 1989.
[6] 池上ほか，" 階層型イベント検知に基づく人と物の関わりのロギングシステム"，第 18 回画像の認識・理解シンポジウム，SS5-37，2015
[7] J. Masci, U. Meier, D. Cirean, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," in Artificial Neural Networks and Machine Learning ICANN 2011, ed. T. Honkela, W. Duch, M. Girolami, and S. Kaski, Lecture Notes in Computer Science, vol.6791, pp.52-59, Springer Berlin Heidelberg, 2011.
[8] P. Baldi and K. Hornik, "Neural networks and principal component analysis: Learning from examples without local minima," Neural Networks, vol.2, no.1, pp.53-58, 1989.
[9] A. Makhzani and B.J. Frey, "k-sparse autoencoders," CoRR, vol.abs/1312.5663, 2013.
[10] I. Ramirez, P. Sprechmann, and G. Sapiro, "Classification and clustering via dictionary learning with structured incoherence and shared features," Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, pp.3501-3508, June 2010.
[11] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," Proceedings of the 26th Annual International Conference on Machine Learning, ICML ' 09, New York, NY, USA, pp.689-696, ACM, 2009.
[12] H. Lee, C. Ekanadham, and A.Y. Ng, "Sparse deep belief net model for visual area v2," in Advances in Neural Information Processing Systems 20, ed. J. Platt, D. Koller, Y. Singer, and S. Roweis, pp.873- 880, Curran Associates, 2008.
[13] D.L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via 1 minimization," Proceedings of the National Academy of Sciences, vol.100, no.5, pp.2197-2202, 2003.
[14] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," Computer Vision and Pattern Recognition, 2009, CVPR 2009, IEEE Conference on, pp.1794-1801, June 2009.
[15]] H. Bristow, A. Eriksson, and S. Lucey, "Fast convolutional sparse coding," Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, pp.391-398, June 2013.
[16] T.Matsuo et.al，"Evaluation Function for Shift Invariant Auto-encoder"，MPR2016，P1-9，2016