

人間の手の姿勢の転写に基づく 多指ロボットの物体把持

福原 宏弥[†] 川上 拓也^{††} 矢野 将基^{††} 松尾 直志^{††} 島田 伸敬^{††}

[†] 立命館大学大学院情報理工学研究科 〒525-8577 滋賀県草津市野路東 1-1-1

E-mail: †fukuhara@i.ci.ritsumeai.ac.jp

あらまし 本研究では想起した人間の手の距離画像を入力として3本指×3関節のロボットハンドに同じ機能をもつ姿勢をとらせることでロボットによる物体把持を行う。Shift Invariant Auto Encoder を用いて、手の距離画像を低次元表現する部分空間を作成する。作成した低次元空間に少数の手の距離画像を投影し、その手の姿勢と同じ機能をもつロボットハンドの姿勢（関節角度）を教師姿勢として与え、回帰モデルを作成した。実際に作成した低次元空間に手の距離画像を投影し、投影したものを回帰モデルに入力し、推定された結果をロボットハンドに姿勢をとらせた。

キーワード Shift Invariant Auto Encoder, ハンドアームロボット, 手の模倣, 半教師あり学習

Object grasping by multi-fingered robot based on transcription of human hand shape

Hiroya FUKUHARA[†], Takuya KAWAKAMI^{††}, Masaki YANO^{††}, Tadashi MATSUO^{††}, and
Nobutaka SHIMADA^{††}

[†] Graduate School of Information Science and Engineering Ritsumeikan University Nojihigashi 1-1-1

Kusatsu-shi 525-8577 Japan

E-mail: †fukuhara@i.ci.ritsumeai.ac.jp

Abstract In this research, we use the depth image of the human hand recalled as input and wake the robot hand of 3 fingers × 3 joints take the posture having the same function to grasp the object by the robot. Using Shift Invariant Auto Encoder, we create a subspace that expresses the depth image of the hand in a sub dimension. A small number of hand depth images are projected in the created sub dimensional space, and a regression model is created by associating the posture (joint angle) of the robot hand having the same function as the posture of the hand. The depth image of the hand was projected to the actually created sub dimensional space, the projected one was input to the regression model, and the estimated result was taken as the posture of the robot hand.

Key words Shift Invariant Auto Encoder, Hand-Arm Robot, Imitation of hand, Semi-supervised learning

1. ま え が き

近年、少子高齢化に伴う労働力の不足から介護分野などでサービスロボットに期待が寄せられている。ロボットが一般家庭などで活躍するには人間と同じように物体を扱う必要がある。これらの物体に合わせて物体を扱うプロセスを人間が考え、ソフトウェアを開発するためには専門的な知識が必要となるうえ、手の動作や姿勢は数多くあり、これらの動作を一つ一つ開発するのは現実的ではない。そこで、ロボットが人間の姿を見て、ロボットが動作を生成することができるようになれば、開発者の負担を軽減することができるようになる。また、人間の

手に基づいてロボットが動作を学習することで、人間の経験に基づく洗練された動作を身につけることができるようになると思われる。

人間の手を観察や参考にしてロボットに模倣させている研究は、デバイスを人間の手に付け、人間の手の関節角度の情報を取得してロボットハンドに動作をさせている研究や、RGBカメラで人間の動作をリアルタイムで取得を行い、人間の動作からロボットの動作を生成する研究などがある。Geng [1] は ShapeHand というモーションキャプチャーモジュールを装着し手の動作をキャプチャーする。その後、物体を3本指で掴み、掴んだときの関節角度と物体の特徴のデータを集め、

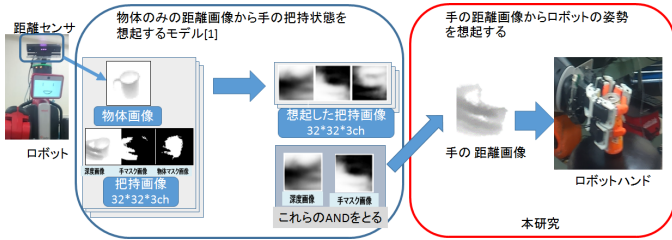


図 1 本研究のコンセプト

GNG(Growing Neural Gas) で学習させる。学習していない物体をロボットに把持をさせるためには手にデバイスを付け、データをとる必要がある。

我々は川上ら [2] の物体形状からの把持パターンの想起を用いて物体把持を行う方法を提案する (図 1)。物体画像からその物体を把持するための人間の手の距離画像が想起する。図 1 の想起した把持画像の距離画像、手マスク画像、物体マスク画像である。手の距離画像から手と同じ機能を持つロボットハンドの姿勢に生成することができれば、物体把持ができる。本研究は人間の手の距離画像から必ずしも人と同じ形状ではないロボットハンドに人と同じ姿勢をとらせて物体把持を行う。

本研究で使用するロボットは Rethink Robotics 社が開発した 7 自由度の双腕ロボット, Baxter を使用する。Baxter の手先には本研究で製作したロボットハンドを装着し, Baxter の頭部に RGB-D センサとして KinectV2 を設置する。

2. 本研究の提案手法

握る, つまむ, 物体を持つ姿勢など人間の手は様々な姿勢をとる。これらの姿勢を手と同じ機能をもつロボットハンドの姿勢と対応付け, CNN(Convolutional Neural Network) の教師付き学習することで回帰モデルを生成する。この方法では人間の手の姿勢とロボットハンドの姿勢をすべて対応付けているので, 人間の手からロボットハンドの姿勢を推定することができる。しかし, 人間の手は様々な姿勢をとり, すべてを考慮してロボットハンドの姿勢と対応付けるのは容易ではない。そこで我々は半教師あり学習で手の距離画像からロボットハンドの姿勢を推定する。まず, 人間の手の姿勢を距離画像で取得し, 手の距離画像を低次元表現する部分空間を生成する。生成した低次元空間に少数の手の距離画像を投影し, その手の姿勢と同じ機能をもつロボットハンドの姿勢 (関節角度) を対応付け, 回帰モデルを生成する。部分空間を生成することで, 部分空間を生成する際に使用していない手の距離画像を入力した際に入力した姿勢と似た姿勢の近くに投影されることを狙う。本研究では手の距離画像を低次元表現する部分空間は主成分分析と Shift Invariant Auto Encoder を使い, 精度を比較した。回帰モデルは全結合を用いて生成する。

3. 手の形の部分空間の生成

学習に使用するデータは手が写った距離画像でデータ数は 9377 枚である (図 2)。本研究では人間の手の模倣するために

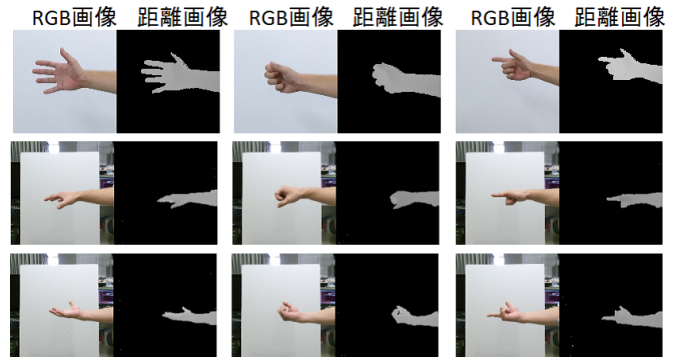


図 2 部分空間の生成に使用した手の距離画像 (一部)

様々な姿勢を集める。上から握む姿勢, 下から握む姿勢, 横から握む姿勢の 3 種類をとる。指の姿勢は筒握り, 指尖つまみ, 指腹つまみ, 鉤握り, 球握り, こぶし握り, 掌側面つまみ, 親指を立てる, 人差し指を立てる, 中指を立てる, チョキの形をした姿勢, 手を開く姿勢 [4] など集め, これらの画像を用いて部分空間の作成を行う。

4. PCA を用いた部分空間の生成

PCA を用いて生成した低次元化した部分空間を検証する。手の距離画像 (図 2) を用いて部分空間を生成する。生成した部分空間に低次元空間の生成に使用していない手の距離画像を投影し復元した結果を評価する。本研究では 100 次元で圧縮を行う。100 次元で圧縮した場合の累積寄与率は 95 % である。復元した結果を図 3 に示す。図 3 の RGB 画像は距離画像を取得した際の様子, 距離画像が PCA で生成した部分空間に投影する画像, 復元画像は入力した画像を復元した結果になる。図 3 は部分空間に生成した手の距離画像と少し異なる画像を入力した結果, どんな指形状の画像でも指が復元されない。姿勢が学習画像と少し異なるものを入力すると復元結果で指など手の形に復元されない。PCA は外乱に弱く, 特に手の領域の画像内シフトに弱い。

5. Shift Invariant Auto Encoder を用いた部分空間の生成

PCA を用いて手の距離画像を低次元で表現する部分空間にテストデータ (図 3) を入力し復元した結果は手の形がぼやける。そこで我々は松尾 [3] が提案した教師無しクラスタリング手法である SIAE(Shift Invariant Auto Encoder) を用いて手の距離画像の部分空間を生成する。SIAE はシフトした画像を入力しても似た形状であれば似た記述子が出力される。つまり, シフトして重なるものは, 同じものになるように集約された空間と写像が得られる。

本研究で用いる SIAE でのネットワーク層は図 4 に示す。入力される画像は 50×50 のサイズの画像から 32×32 の画像を切り取り, バッチ画像を生成する。学習に使用する画像のサイズは 32×32 pixel の 1024 ユニットを入力とする。画像の特徴を抽出するために特徴量を自動で取得する畳み込みとプーリン

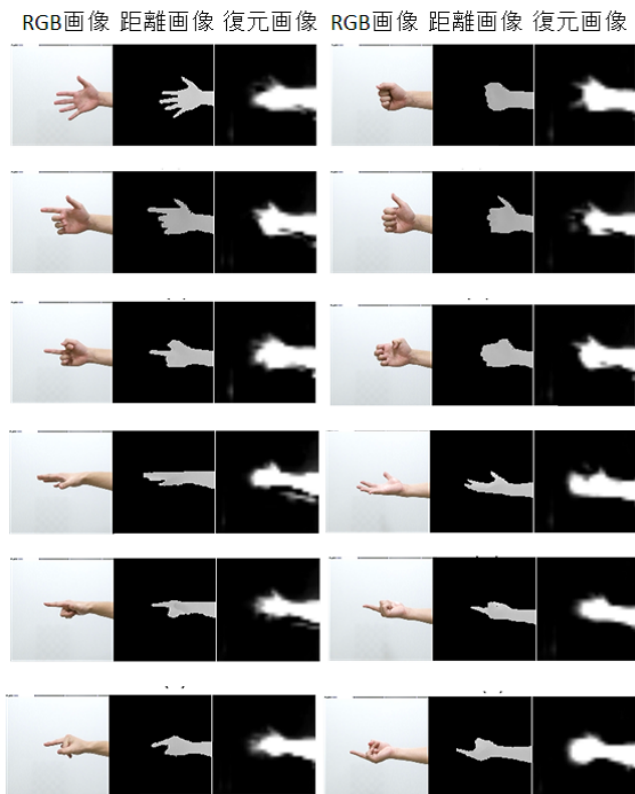


図3 テスト画像を投影して復元した結果 RGB 画像:手の距離画像を取得した際の RGB 画像 距離画像:生成した部分空間に投影する距離画像 復元画像:投影後に復元した結果

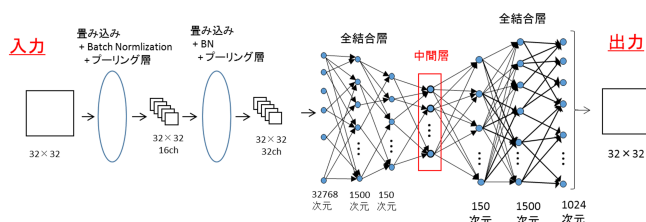


図4 Shift Invariant Auto Encoder のネットワーク構成図

グの処理を2層、構築しエンコーダまでの全結合の層を3層とする。デコーダの復元部の全結合の層は3層とした。エンコーダは入力した手の距離画像を200次元に圧縮し、デコーダは入力した手の距離画像を復元するようにする。学習したモデルを用いて精度の確認を行う。

生成したモデルに手の距離画像を入力しエンコード・デコードをした結果を図5に示す。図5のRGB画像が手の距離画像を取得した際のRGB画像、距離画像が生成したモデルに入力する距離画像、SIAEが投影後、復元した結果、PCAで投影後、復元した結果である。図5の(a)の手が開いた姿勢の復元結果を見るとSIAEは5本の指、すべて復元されていることがわかる。また、図5の(b)の人差し指と親指を立てる姿勢でも、手の形が復元されていることがわかる。この結果からSIAEのほうが手の距離画像の記述力が高いと言える。SIAEで生成した部分空間を用いて回帰モデルを生成する。

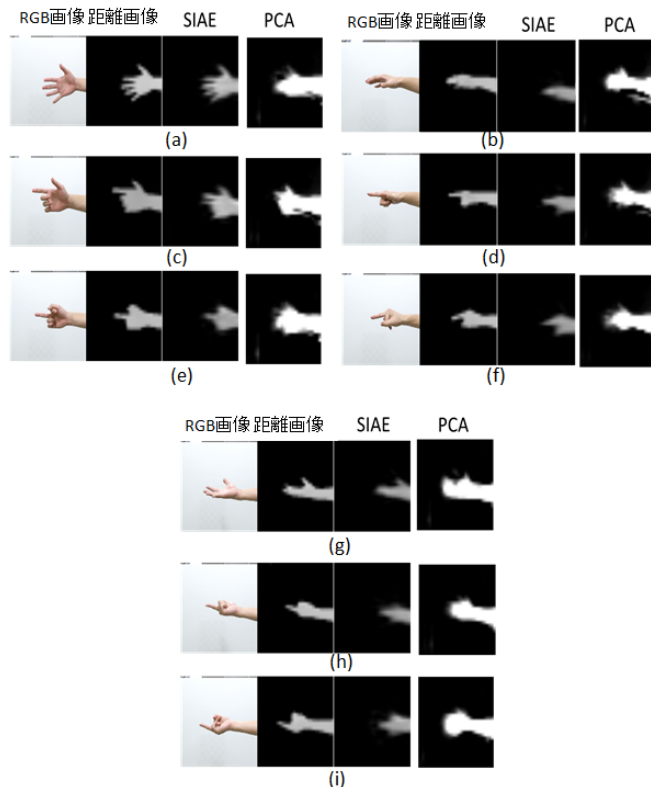


図5 テスト画像をモデルに入力しエンコード・デコードをした結果 RGB 画像:手の距離画像を取得した際の RGB 画像 距離画像:生成したモデルに入力する距離画像 SIAE:SIAEで投影後に復元した結果 PCA:PCAで投影後に復元した結果

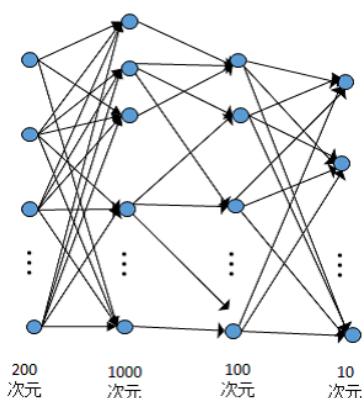


図6 回帰モデルを作成するためのネットワークの構成

6. 手の距離画像からのロボットハンドの姿勢

手の距離画像のエンコーディングとロボットハンドの姿勢の回帰モデルの生成について説明する。全結合のネットワーク層は図6になる。ネットワークの深さは4層に設定した。入力手の距離画像をエンコードした200次元のベクトルになる。出力するものはロボットハンドの9つの関節とBaxterのW2の関節である。出力される次元数はロボットハンドの9つの関節とBaxterのW2の関節で10次元になる。

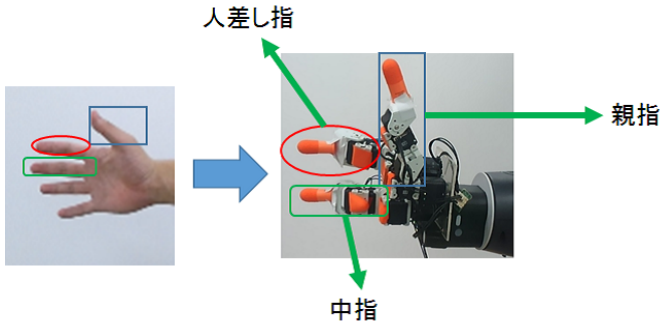


図 7 ロボットハンドの指の割り当て

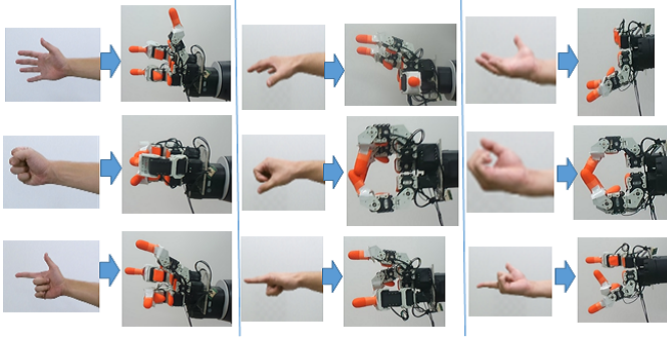


図 8 人間の手とロボットハンドの対応付け 左の画像:手の姿勢, 右の画像:対応付けるロボットハンドの姿勢

6.1 低次元化した人間の手の距離画像とその手の姿勢と同じ機能を持つロボットハンドの姿勢の対応付け

人間の手が写った距離画像とロボットハンドの姿勢 (関節角度) の対応付けについて説明をする。本研究で使用するロボットハンドは 3 本指しかないで人間の手と同じ形をした姿勢はできない。ロボットハンドの 3 本指を人間の 5 本指のどれかに対応付ける必要がある。人間の手の親指, 人差し指, 中指を物体操作に使っていることから, この指をロボットハンドの 3 本指に対応付ける (図 7)。ロボットハンドと人間の右手の姿勢を対応付けて, ロボットハンドのとりべき手首と指の教師姿勢を与えた。握る, 親指を立てる姿勢などをそれぞれロボットハンドで再現し, 教師姿勢として対応付けた人間の手の姿勢とロボットハンドの姿勢を図 8 に示す。人間の手の姿勢とロボットハンドの姿勢の対応関係は必ずしも類似していなくても構わないが, ここでは同じ把持タイプを実現する姿勢を選んで教師姿勢を与えた。手首の動きは双腕ロボット Baxter の関節 W2 が人間の手首にあたるので, W2 (図 9) を回転させて表現した。W2 の関節が 0 度では手が横から物を掴む姿勢を示し, 90 度では手が上から物を掴む姿勢に, -90 度では手が下から物を掴む姿勢を示す。対応付けた姿勢の数は 1339 個である。

6.2 回帰モデルの精度評価

図 10 は PCA と SIAE のエンコーディングとハンドの姿勢の回帰モデルをそれぞれ生成したものと, 部分空間を作成せず CNN で手の距離画像とロボットハンドの姿勢の回帰モデルを生成し, それぞれロボットハンドの姿勢を推定した結果である。CNN の学習データ数は PCA と SIAE で対応付けた 1339 個と

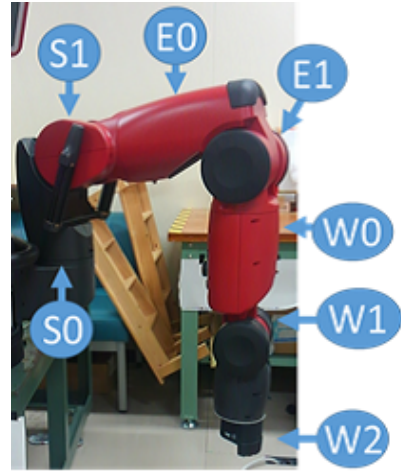


図 9 双腕ロボット Baxter の関節名

同じである。推定する手の距離画像は学習に使用していないテストデータを用いた。

図 10(a) は手が開いた姿勢を入力したもので, SIAE と CNN を比較したとき, 指の姿勢は似た姿勢をとっているが CNN は手首の姿勢により大きな誤差がでていることがわかる。また, 表 1 はそれぞれのロボットハンドとロボットの手首の関節角度の誤差の RMSE (root mean squared error) である。この教師データは推定する手の姿勢と学習データで似た姿勢に対応付けたロボットハンドの姿勢とした。表 1 から SIAE が最も誤差が小さいことがわかる。図 10(b) は手を広げ上から掴む姿勢の推定結果である。表 1 をみると SIAE のほうが精度がよいことがわかる。図 10(c) は握る姿勢の推定結果で SIAE と CNN はほぼ似た形の姿勢をとっているが CNN はロボットハンドの親指が SIAE と比べ誤差が大きい。図 10(d)(e)(f) は CNN では手首の推定で大きな誤差がでている。それと比べ SIAE は誤差が小さい。PCA でエンコードした記述子から推定したものは, 図 10(a) から (f) すべて入力した姿勢と似ていない姿勢になっている。手の距離画像のテストデータ 27 個を用いて, SIAE でエンコードした記述子を回帰モデルに入力し推定した各関節角度の誤差と CNN で推定した各関節角度の誤差のヒストグラムを図 11 に示す。図 11 の左の 5 つのヒストグラムは SIAE でエンコードした記述子から推定した各関節角度の誤差のヒストグラム, 右は CNN で推定した各関節角度の誤差のヒストグラムとそれぞれの関節角度の RMSE を示す。図 11 の上から手首の関節角度, 親指の関節角度, 人差し指の関節角度, 中指の関節角度, 手首以外のロボットハンドの関節角度のそれぞれの誤差のヒストグラムを示している。図 11 の各関節角度の RMSE では CNN の方がよいことがわかる。しかし, 図 11 のヒストグラムを見ると SIAE の方が誤差の小さいサンプル数が多いことが分かる。特に手首の関節角度の誤差のヒストグラムは誤差が小さいものが多い。RMSE が大きくなっているのは -170 など大きい値が出力されたため RMSE が大きくなった。親指, 人差し指, 中指に関しては SIAE と CNN での RMSE は大きな違いはなくヒストグラムもあまり大きな違いがない。手首以外のロ

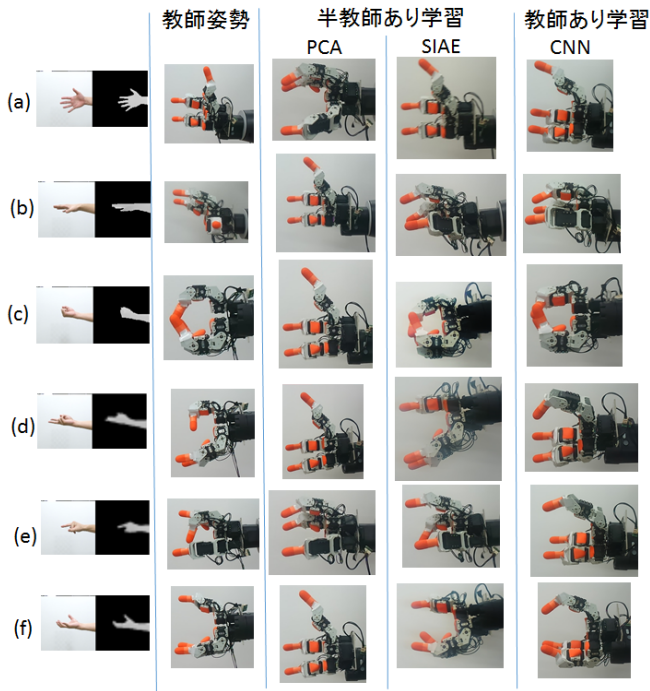


図 10 半教師あり学習 (PCA, SIAE), 教師あり学習 (CNN) を用いてロボットハンドの姿勢を推定した結果

表 1 教師データと推定値の誤差の RMSE(単位:度)

図番号	PCA	SIAE	CNN
(a)	34.5	12.5	39.6
(b)	24.5	18.0	22.5
(c)	55.8	9.6	17.0
(d)	46.2	19.2	25.2
(e)	21.3	15.2	37.2
(f)	46.2	16.0	19.7

ロボットハンド (9 つの関節) の RMSE を見ると違いはあまりないが、ヒストグラムを見ると SIAE の方が 0 度付近に分布しているサンプル数が多いことが分かる。SIAE で半教師あり学習が他の 2 つと比べると精度がよいことがわかった。

7. 物体のみの距離画像から手の把持状態を想起するモデルを用いた物体把持

図 12 に持ち方を表すパラメータを想起する把持パターン想起モデルを作成する手順を示す。SIAE を用いて把持画像から 30 次元の持ち方パラメータ空間を学習する。持ち方パラメータの作成にエンコーダ部を用い、持ち方パラメータから把持画像を復元するにはデコーダ部を利用する。次に学習済み SIAE 持ち方パラメータを教師とし、CNN で物体のみ画像との関係を学習する。

図 13(a) は取っ手があるコップの距離画像、(b) は取っ手のないコップの距離画像で (c) はボールの距離画像である。図 14(a)(b)(c) はそれぞれ図 13 の把持画像を把持パターン想起モデルに入力して想起された画像である。図 14 の手のマスク画像で指の形が想起されていることが分かる。また、距離画像で手と物体の形が想起されていることから期待通りの結果を得られ

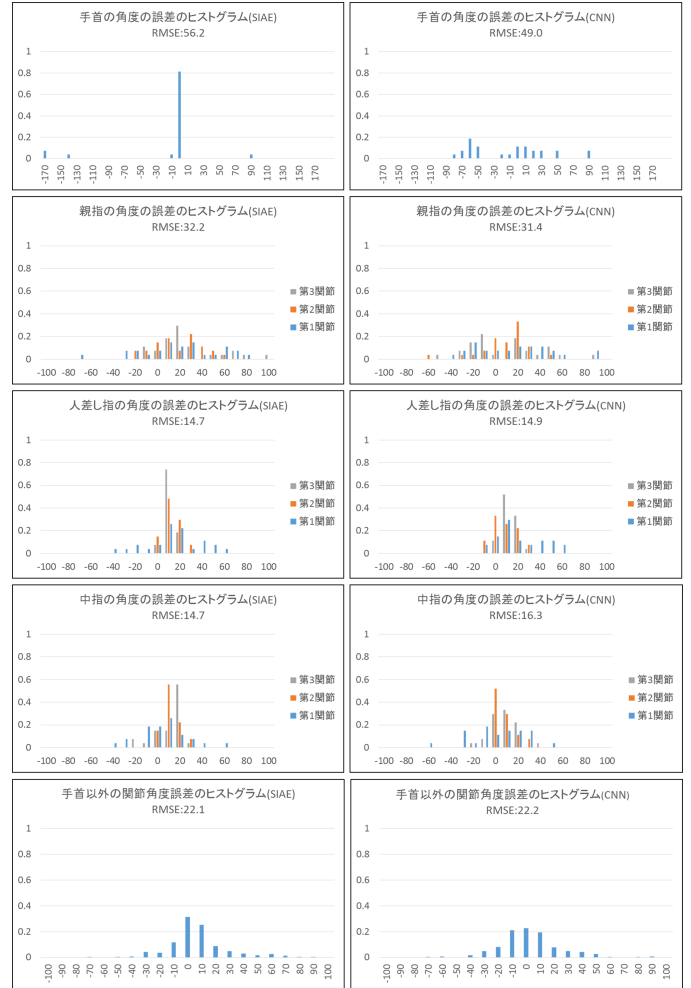


図 11 SIAE でエンコードした記述子を回帰モデルに入力し推定した各関節角度の誤差と CNN から推定した各関節角度の誤差のヒストグラム (左のヒストグラム: SIAE 右: CNN のヒストグラム 上から手首の関節角度, 親指の関節角度, 人差し指の関節角度, 中指の関節角度, 手首以外のロボットハンド (9 つの関節) の関節角度のそれぞれの誤差のヒストグラム)

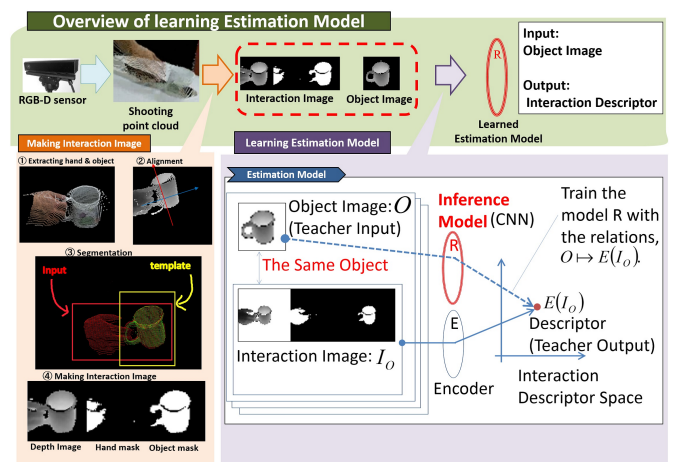


図 12 把持パターン想起モデルの作成手順

ている。図 14 の手のマスク画像と距離画像を用いて距離画像からそれぞれ手の領域を切り取った画像が図 15 である。部分空間の生成で用いた学習画像と図 15 の手の写り方が大きく異

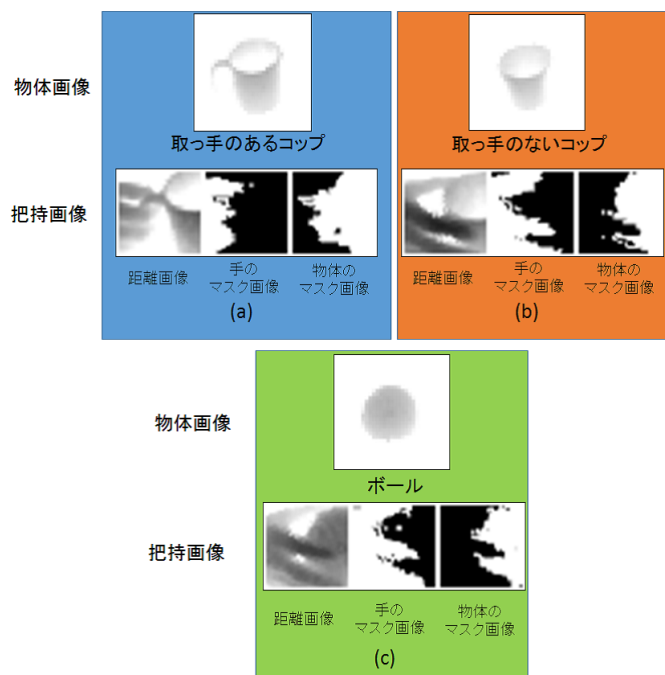


図 13 把持パターン想起モデルに入力する把持画像

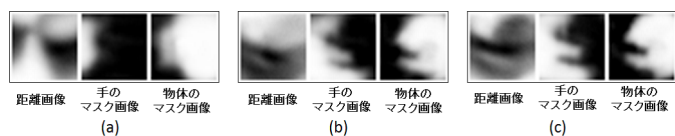


図 14 把持パターン想起モデルから想起された把持画像

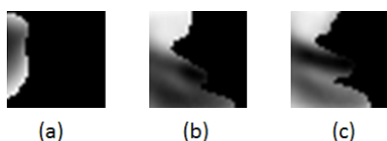


図 15 想起された距離画像から手のマスク画像を用いて切り取った画像

なっている。図 15 の画像を本研究で作成した SIAE の部分空間に投影しても、期待した結果は得られない。しかし、6.2 節の結果から、図 15 のような画像を集め、SIAE で部分空間を生成し生成した部分空間に手の距離画像を投影して、その手と同じ機能をするロボットハンドの姿勢を教師姿勢として与え回帰モデルを生成すれば、ロボットハンドの姿勢を推定することができる (図 16)。

8. まとめと今後の課題

本研究では Shift Invariant AutoEncoder を用いて手の距離画像の部分空間を生成をし、その部分空間のサンプル点のいくつかとロボットハンドの姿勢を対応付け、回帰モデルを作成することで、10 次元のロボットの関節の推定をすることができた。

今後の課題は把持パターン想起モデルに使用した手の姿勢の学習を行い部分空間を作成し、手の距離画像とロボットハンドの姿勢の対応付けを行い、ロボットハンドの姿勢の推定する。推定した結果をロボットハンドに姿勢をとらせ、入力した姿勢と

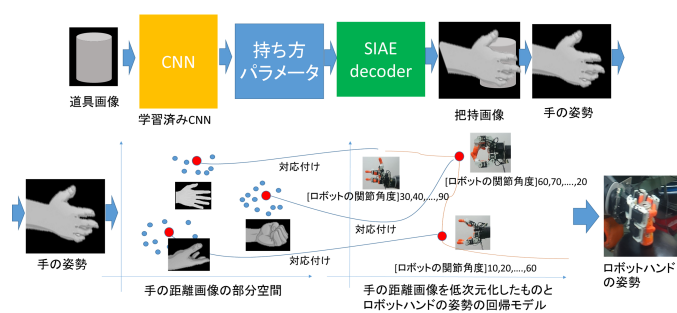


図 16 把持パターン想起モデルと本研究で作成したモデルを用いたロボットハンドの物体把持 (イメージ)

似た姿勢をとるか実験を行う。また、把持パターン想起モデルから把持画像の想起結果を用いて、ロボットの姿勢を想起し、想起した姿勢を初期値としてロボットが物を掴む制御を実装する。実際に物がロボットハンドに触れた際に力を入れ続ける制御を入れ、物体を持ち上げる実験をする。

部分空間の記述の精度をあげるために、RGB 画像を用いて、手のテクスチャを獲得し、学習を行う。手の距離画像の手の形のみでは手首の向きを推定することが難しい姿勢がある。例えば、手を握る姿勢である。握った姿勢で手首を回転させても手の形に大きな違いがない。手のテクスチャがあれば、握った姿勢でも違いがあるので、手首の向きがわかる。そのため、ロボットハンドの姿勢の推定精度が向上する。

文 献

- [1] Geng, Tao, et al. "Synergy-based affordance learning for robotic grasping. " Robotics and Autonomous Systems 61. 12 (2013): 1626-1640.
- [2] 川上拓也 松尾 直志 小川陽子 島田伸敬 "3-D シーン観察に基づく手と物体の関係性の学習と把持パターンの想起" 電子情報通信学会の ISS 特別企画, 学生ポスターセッション (2016): ISS-SP-195.
- [3] T. Matsuo et. al, " Evaluation Function for Shift Invariant Auto-encoder ", MPR2016, P1-9, 2016
- [4] Noriko Kamakura: "手の形 手の動き", 医歯薬出版株式会社, 1989.