

RGB-D 画像からの把持パターン想起に基づくハンドアームによる物体把持

Grasping an Object by a Hand-Arm Robot Based on Human Interaction Recalled from RGB-D Image

○ 矢野将基 (立命館大) 福原宏弥 (立命館大)
正 松尾直志 (立命館大) 正 島田伸敬 (立命館大)

Masaki YANO, Ritsumeikan University
Hiroya FUKUHARA, Ritsumeikan University
Tadashi MATSUO, Ritsumeikan University
Nobutaka SHIMADA, Ritsumeikan University

We propose a method that enable a robot to grasp on object based on how a human grasp it. By observing interactions by humans, we model the relationship between a shape of an object and how to grasp it. In advance, the model is trained with pairs of images before/after a human grasps an object. We can train the model without labeling interaction between an object and a hand. With this model, a robot can recall how a human grasps an object from an appearance of the object. The robot can grasp the object by moving its hand to a point of the object where a human touches for grasp it. By experiments for actual objects, we show availability of proposed technique.

Key Words: Robot, Deep-Learning, Grasp Object

1 はじめに

近年、カメラやセンサの高性能、低価格化が進んでおり、高性能なセンサをロボットに搭載することが可能となっている。ロボットは一般家庭など、複雑な環境下にも活躍の幅を広げていくと考えられる。しかし、ロボットの動作一つ一つのソフトウェアを人間が開発するのは開発者への負担が大きくなるため困難である。そこで、人間が道具を操作するシーンをロボットが観察し、操作を自動的に学習することができれば、この問題の解決につながると思われる。

図1に提案する手法の概要を示す。本研究では、人間が物体を把持した様子を学習し、把持パターンを自動で取得する手法を提案する。物体に対する人間の把持パターンをあらかじめ学習しておく、物体が写った画像が与えられた際に把持パターンを想起する。そして、想起結果から手先の目標位置と物体の中心位置を取得し、ハンドアームロボットによって物体把持を行う。

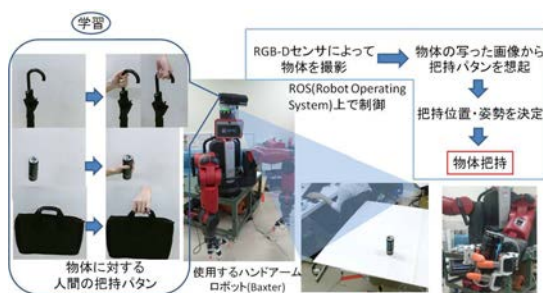


Fig.1 提案する手法の概要

本研究では、Rethink Robotics 社が開発・販売している7自由度の双腕ロボット、Baxter を使用する。Baxter の手先には本研究室で製作したロボットハンドを装着し、Baxter 頭部には RGB-D センサとして Kinect V2 を設置する。

2 RGB-D 画像からの把持パターンの想起

人間が物体を把持するときの手の姿勢をその物体に対する把持パターンという。鎌倉 [1] によると、人間は物体の使用目的に応じ

て把持パターンを変えているといい、その種類は限定されているという。これより、同じ使用目的の物体ならばほぼ同じ把持パターンになると考えられる。

物体のみが映った画像から把持パターンを推定することを把持パターンの想起とする。松尾ら [2] はある物体の典型的な把持パターンを学習することで、未知の物体に対する把持パターンを想起する手法を提案している。この手法では、CAE(Convolutional Auto-Encoders) を用いて道具ごとの持ち方を表す持ち方パラメータを抽出し、それを入力として未知の物体画像から把持パターンを想起するための学習を行う。

把持パターン想起は、濃淡画像を入力とし、入力画像の各位置ごとに 32×32 のウィンドウ(パッチ)を生成する。そして、パッチごとに持ち方パラメータと手と物体の相互作用が起きている確率の想起を行い(図2)、把持パターンの想起を行う(図3)最後に、手と物体の相互作用確率を重みとしてパッチごとの想起結果を足し合わせ、その結果画像を出力とする。

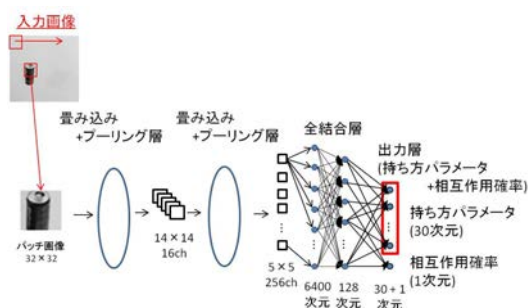


Fig.2 パッチごとの持ち方パラメータ想起

今回は想起結果のうち手領域マスク画像と物体領域マスク画像を使用して手先の目標位置と物体の中心位置を推定し、物体把持を行う。

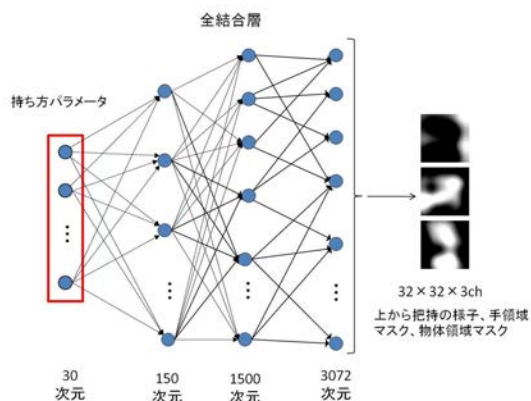


Fig.3 持ち方パラメータからの把持パターン想起

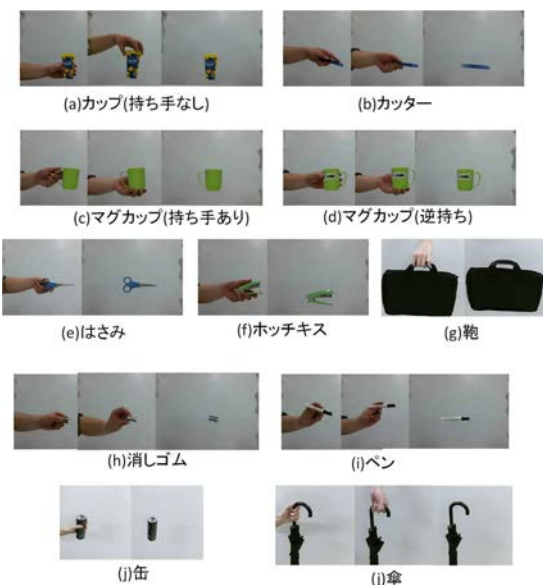


Fig.4 学習に使用する物体と把持パターン

3 想起のための手と物体の関係性の学習と学習結果の評価

学習に使用する画像は、図4に示した通りの持ち方で、持ち方の数は計18通りである。各持ち方に対する画像の枚数は120枚であり、 $120 \times 18 = 2160$ 枚の画像を使用して学習を行う。

図5は学習に使用した缶(図4の(j))について把持パターンの想起を行った結果である。ここで、図5(e)の入出力を重ね合わせた画像は、青色部分は把持の様子を表すテキストチャ、緑色部分は手領域のマスク、赤色部分は物体領域のマスクを表している。図5の(e)と(f)を比較すると、(f)における缶の領域は赤色、手領域は緑色、その周辺は青色に塗り分けられている。なお、図5(e)の手首位置より左の部分は緑色に塗られていない。理由として、今回は 32×32 のウィンドウ毎に想起を行っており、手首付近のウィンドウに缶が入らなかったためであると考えられる。このことから、学習済み画像について概ね期待通りの結果が得られていると考えられる。

4 想起結果に基づく把持動作の実装

4.1 RGB-D 画像からの把持パターン想起

Kinect V2によって取得したカラー画像には背景も含まれているが、今回は手先の可動範囲外の領域については考慮する必要がない。そこで、Kinect V2によって取得した深度情報を使用し、想起領域の抽出を行う。本研究では距離が1.5m以内の領域に対してラベリング処理を行い、領域数が最大となる領域に対して想起を行う。

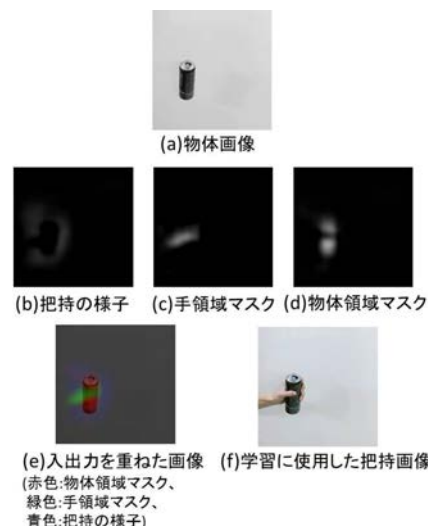


Fig.5 学習に使用した物体に対する把持パターン想起結果

想起結果からロボット座標系における手先の目標位置と物体の中心位置を取得する。まず、想起結果画像のうち手先領域マスク画像と物体領域マスク画像に対して、しきい値を基に2値化処理を行う。そして、2値化画像に対してラベリング処理を行い、領域数が最大となるラベルの重心を手先の目標位置・物体の中心位置の座標とする。

次に、求めた座標を画像座標系からカメラ座標系に変換し、さらにロボット座標系に変換する。画像座標系からカメラ座標系への変換には本研究室内で制作したKinect座標系変換サーバを使用する。座標系変換サーバではKinect SDKに用意されているCoordinateMapperクラスのメンバ関数である、MapColorFrameToCameraSpace関数を使用して座標系の変換を行っている。カメラ座標系からロボット座標系への変換は、キャリブレーションによって変換行列を作成し、この変換行列によって行う。カメラ座標系のある点 (X_{ci}, Y_{ci}, Z_{ci}) と、それに対応するロボット座標系の点 (X_{ri}, Y_{ri}, Z_{ri}) の組を10組程度を取得し、疑似逆行列を計算することで変換行列を求めることができる。

4.2 想起結果に基づく手先移動と把持

手先の目標位置と物体中心位置より把持の際の手先姿勢を求める。手先姿勢は図6に示す通り、手先の目標位置から物体中心位置に向かうベクトルと手先の法線ベクトルが一致する姿勢とする。今回はロボットアームの制御に使用するため、目標となる手先姿勢を四元数(quaternion)の形で求める。

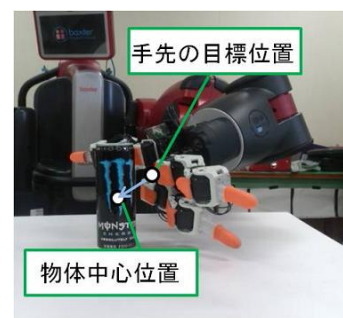


Fig.6 手先の目標姿勢

手先の目標位置 (X_g, Y_g, Z_g) から物体中心位置 (X_o, Y_o, Z_o) へのベクトルと直交するベクトルのうち、Z成分が0(=水平)となるベクトルを回転軸ベクトル $V(V_x, V_y, V_z)$ とすると、回転軸ベクトルと回転角度 θ は以下の式で表わされる。

$$V_x = -\sqrt{(X_o - X_g)^2 + (Y_o - Y_g)^2 + (Z_o - Z_g)^2} * (Y_o - Y_g) \quad (1)$$

$$V_y = \sqrt{(X_o - X_g)^2 + (Y_o - Y_g)^2 + (Z_o - Z_g)^2} * (X_o - X_g) \quad (2)$$

$$V_z = 0 \quad (3)$$

$$\theta = \arccos\left(\frac{Z_o - Z_g}{\sqrt{(X_o - X_g)^2 + (Y_o - Y_g)^2 + (Z_o - Z_g)^2}}\right) \quad (4)$$

回転軸ベクトル V と回転角度 θ から目標手先姿勢を表す四元数 $Q(Q_x, Q_y, Q_z, Q_w)$ を求めると、以下の式で表される。

$$Q(Q_x, Q_y, Q_z, Q_w) = \left(V_x \sin\left(\frac{\theta}{2}\right), V_y \sin\left(\frac{\theta}{2}\right), V_z \sin\left(\frac{\theta}{2}\right), \cos\left(\frac{\theta}{2}\right)\right) \quad (5)$$

手先の目標位置と姿勢から逆運動学を計算することで、各関節の目標値を取得し、制御を行う。ロボットハンドの把持姿勢として缶を把持した状態(図7)の関節角度を記録しておく。そして、手先が目標に移動した後にその角度を再現することで把持を行う。

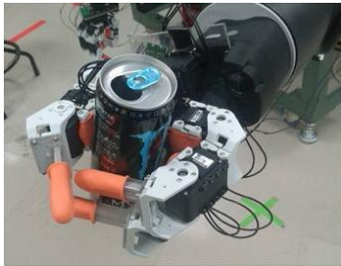


Fig.7 缶を把持した状態

5 物体把持実験

これまで述べてきた手法により、実際にロボットが物体を把持できるかどうかを検証する。

5.1 実験環境

今回は、図8に示す缶とカバンに対して把持実験を行う。これらはいずれも学習に使用したものである。



(a) 実験に使用する缶 (b) 実験に使用するカバン

Fig.8 実験に使用する物体

また、実験に使用する物体は図9に示す通り、いずれも高さ70cmの台上、70cm先の位置に立てて配置した。

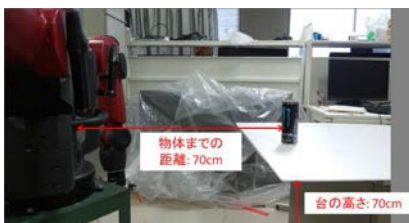


Fig.9 実験環境

5.2 缶を把持する場合の実験結果

5.2.1 RGB-D 画像からの把持パターン想起

図10はロボット頭部に設置した Kinect V2 によって取得したカラー画像である。Kinect V2 によって取得した深度情報を使用し、図10のうち、距離が1.5m以内の領域を抽出した画像を生成し、その画像に対して把持パタンの想起を行う。



Fig.10 Kinect V2 により取得したカラー画像

サーバへの入出力画像を重ね合わせた結果を図11に示す。赤色が物体領域、緑色が手領域を表すマスクである。先述の通りウィンドウごとの想起結果を確率に基づいて足し合わせているため、色が濃い部分ほどそれぞれの領域である確率が高いことを示している。缶の上部を中心に物体領域、その左に手領域を表すマスクが示されている。また、右上の台と背景の輪郭部分にも示されている。



Fig.11 入力画像と想起結果を重ね合わせた様子; 赤色:物体領域マスク、緑色:手領域マスク

想起結果の手領域と物体領域マスク画像に関して、しきい値を基に2値化処理を行う。2値化画像を図12に示す。本実験では手領域マスクのしきい値を0.10、物体領域マスクのしきい値を0.15に設定した。2値化処理によって得られた領域に対してラベリング処理を行い、領域数最大ラベルの重心位置を求めて手先の目標位置・物体中心位置とした。図12より、缶の上部に対して手先の目標位置・物体中心位置が示されている。

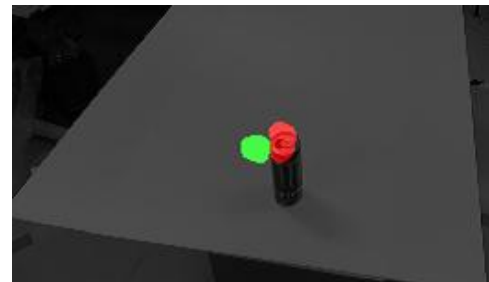


Fig.12 想起結果のうち、物体領域・手領域に関してしきい値を超えた領域; 赤色:物体領域マスク、緑色:手領域マスク

5.2.2 想起結果に基づく手先移動と把持

図13に把持動作開始時からの動作の様子を示す。図12で示した通り、缶の上部を把持している。今回の実験では、開始位置から目標位置に移動するまでが約9秒、目標位置に移動してから把持動作を行うまでが約1.3秒であった。最終的に、缶を倒すことなく把持することができた。

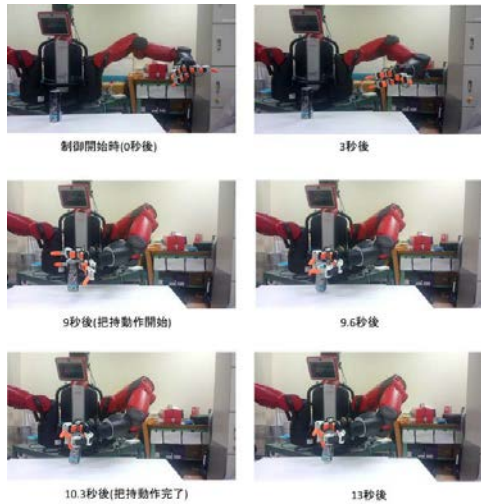


Fig.13 缶を把持する様子

5.3 カバンを把持する場合の実験結果

5.3.1 RGB-D 画像からの把持パタン想起

図 14 は Kinect V2 によって取得したカラー画像である。カバンについても同様に、図 14 のうち、距離が 1.5m 以内の領域を抽出した画像を生成し、把持パタンの想起を行う。



Fig.14 Kinect V2 により取得したカラー画像

サーバへの入出力画像を重ね合わせた結果を図 15 に示す。物体領域はカバンの左端上部と中央から右端上部、手領域は持ち手の上部にマスクが強く示されている。



Fig.15 入力画像と想起結果を重ね合わせた様子; 赤色:物体領域マスク、緑色:手領域マスク

想起結果の手領域と物体領域マスク画像に関して、しきい値を基に 2 値化処理を行う。2 値化画像を図 16 に示す。本実験では手領域マスクのしきい値を 0.10、物体領域マスクのしきい値を 0.15 に設定した。図 16 より、物体領域はカバンの左端上部と中央から右端上部、手領域は持ち手の上部に現れた。物体位置は領域数が大きい方の重心位置とするため、中央から右端上部の物体領域の重心位置を物体の中心としている。

5.3.2 想起結果に基づく手先移動と把持

図 17 に物体付近の動作開始時からの様子を示す。今回の実験では、開始位置から目標位置に移動するまでが約 7.6 秒、目標位

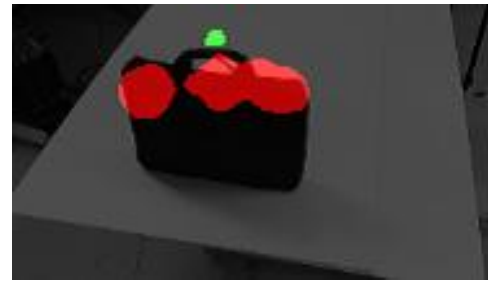


Fig.16 想起結果のうち、物体領域・手領域に関してしきい値を超えた領域; 赤色:物体領域マスク、緑色:手領域マスク

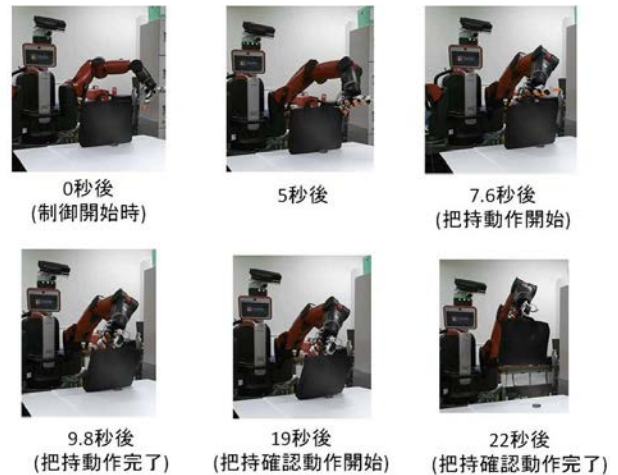


Fig.17 カバンを把持する様子 (物体付近を拡大)

置に移動してから把持動作を行うまでが約 2.2 秒であった。また、カバンは把持後も机上に接着していたため、把持できているかどうかを確認するために腕を上げるための制御コマンドを手動で送信し、確認を行った。確認のための制御コマンドは制御開始から約 19 秒後に送信した。腕を上げた後も、カバンを落とすことなく把持することができた。

6 おわりに

本研究では RGB-D 画像から把持パタン想起を行い、想起情報を用いて物体把持を行う手法を提案した。そして缶について把持実験を行い、手法の有用性を検証した。実験では、缶の下部において物体・手領域マスクが期待通りに示されなかった。原因としては机の上に置いた物体の見え方と学習時の見え方が異なっていたことが挙げられる。そのため、異なった見え方の学習画像を追加して学習を行うことで解決できると考えられる。

今後の課題として、3 次元把持パタン想起への拡張が挙げられる。現在、ロボットハンドによる握り方はあらかじめ作成した姿勢を再現しているため、どの物体に対しても同じ握り方で把持することができるようになると、画像から物体の把持姿勢を獲得することができるようになり、開発者が指示を与える必要がなくなり、負担の軽減につながると考えられる。

参考文献

- [1] Noriko Kamakura: "手のかたち手のうごき", 医歯薬出版株式会社, 1989.
- [2] Tadashi Matsuo, Nobutaka Shimada (Ritsumeikan Univ.): "Construction of Latent Descriptor Space of Hand-Object Interaction", The 22nd Joint Workshop on Frontiers of Computer Vision (FCV2016), O5-1, 2016.