

ヒューマノイドロボットの画像認識に基づく 行動知識の獲得への取組み

An Approach to Action Knowledge Acquisition based on
Image Recognition of a Humanoid Robot

恒川英里 *1
Eri Tsunekawa

小林一郎 *1
Ichiro Kobayashi

麻生英樹 *2
Hideki Asoh

持橋大地 *3
Daichi Mochihashi

ムハンマド アッタミミ *4
Muhammad Attamimi

中村 友昭 *4
Tomoaki Nakamura

長井 隆行 *4
Takayuki Nagai

*1 お茶の水女子大学
Ochanomizu University

*2 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology

*3 統計数理研究所
The Institute of Statistical Mathematics

*4 電気通信大学
The University of Electro-Communications

It is necessary for robots to be practically used in the real world, they have to be able to behave properly to the given task according to the information they observe with their sensors. In this study we take a basic study on the acquisition of robot's action knowledge to the given task by means of Q-learning based on the recognition of the visual information obtained from a robot's hand camera. In concrete terms, we use a humanoid robot HIRO made by Kawada Industries, Inc. and make it learn how to collect colored blocks in a predefined correct order by means of Q-learning based on visual information of its hand camera. We show the results of the experiments.

1. はじめに

近い将来、家庭にロボットが導入され高齢者の支援や居住者の生活を支援することが予想される。その際、ロボットが現実世界に適応するために日常生活に起こりうる様々な課題をこなす必要がある。このことから本研究では、ロボットが現実世界において視覚から得た情報を用いることにより、自らの適切な行動を獲得する強化学習の枠組みについて考察する。

具体的な課題として、テーブル上に置かれた物体を色によって、決められた順番に従うように取得するという行動知識をヒューマノイドロボットを使って実現することに取り組む。

2. ロボットの行動知識獲得

2.1 ヒューマノイドロボット HIRO

使用するロボットは(株)川田工業社製ヒューマノイドロボット HIRONXCを用いる。使用した左腕の性能について、関節は腕に6つ、手に4つあり、手の部分の出力トルクは7.1kgfcm、動作スピードは0.11sec/60°である。画像処理に用いたハンドカメラについて、有効画素数が640×480ピクセルである。このカメラを用いて、テーブル上に置かれている色付きの物体の画像を取得し、画像中の物体に対して、画像処理ライブラリOpenCVを用いた色認識および領域抽出による物体の認識を行う。

2.2 行動知識獲得の概要

テーブルの上に置かれた物体の色画像認識結果に基づき、シミュレーションの中でQ学習を行い、物体を予めこちらが決めた順番に5色を正しく並べるといった知識を獲得するというものである。図1に行動知識獲得の概要を示す。

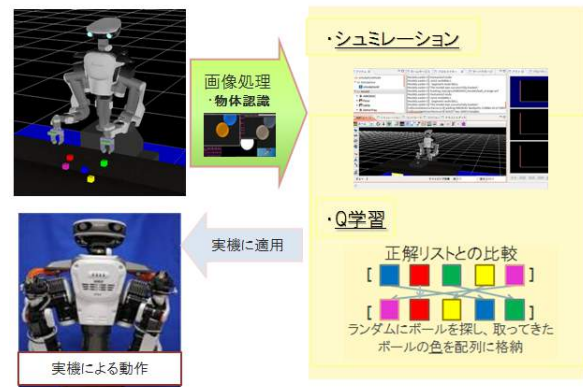


図 1: 行動知識獲得の概要

2.3 画像処理による物体の位置推定

HIROは備え付けられたハンドカメラにより画像の取得を行い、色認識、二値化処理および物体抽出処理を行う(図2)。それにより物体の座標推定を行い、その座標に手を移動し物体の把持を行う。

2.4 強化学習への定式化

2.4.1 Q学習

本研究では、強化学習の枠組みにおいて最適な行動を学習するQ学習[1]を用いることによりHIROの行動知識を獲得する。Q学習による行動価値の更新は、式(1)によって示される。

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha [r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)] \quad (1)$$

上式において、 s は状況、 a は行動、 r_t は時刻 t における報酬、 $Q(s, a)$ は累積報酬 $E\{R_t | s_t = s, a_t = a\}$ で表現される行

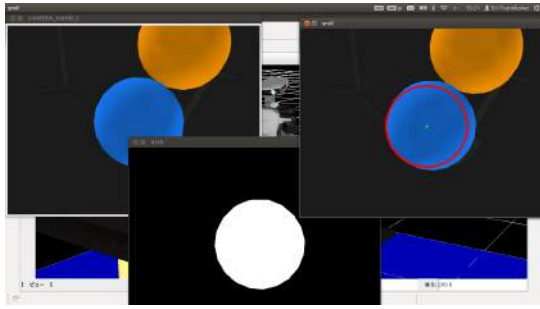


図 2: HIRO による画像処理

動価値を表し, α は学習率, γ は将来の報酬に対する割引率を表す.

2.4.2 状態

実世界の観測に基づく状態の設定に浅田らの先行研究 [2] を参考にする. HIRO が認識する状態は, カメラから得られる視覚情報と物体の取得状態の二つから構成される. ただし, 色によって状態が異なることとするため, 以下に示す 3 つとする.

- 画像中に指定された色の物体が映っている
- 画像中に物体が映っていない
- これまでに取得した物体の順番

2.4.3 行動

HIRO による物体取得の行動においては, まず, カメラに物体が映っていない場合は, 物体を探すために適当な範囲で手を動かす動作が必要になる. また, カメラに物体の一部が映しだされた場合 (本研究では物体を色で認識しているため, 正確にはその色が認識された場合), 物体を把持できる位置に手を動かす動作を行う, 最後にその状態において対象となる物体を取得する/しないの行動が選ばれる. このことから本課題において対象となる動作は以下の 3 つとなる.

- ランダムに手を動かす
- 物体が映ったらその方向に手を動かし, 物体全体を捉える
- 物体を掴む

2.4.4 報酬

互いに異なる色の物体がテーブル上に n 個存在し, それを取ってきた際の報酬は予め決められた順番との差異がペナルティとして与えられる. 上記のペナルティによって物体を正しい順番に整理するための工夫として, 配色に対して異なる価値を付与することを考える. いま, 取得したい順番が [青, 緑, 橙, 白, 黒] とし, リストの右にいく (順番が遅くなる) につれて, その価値が小さくなっているとする. ここでは, 例えば, 価値を青: 5, 緑: 4, 橙: 3, 白: 2, 黒: 1 と設定する. 評価は物体が取得される毎に行われ, それぞれの試行において取得した物体の色と正解順番との差分をペナルティして掛け, 報酬とする. 例えば, [黒, 緑, 青, 橙, 白] という順番を取得した時その時刻の, 報酬は $5 \times (-2), 4 \times 0, 3 \times (-1), 2 \times (-1), 1 \times (-1)$ として与えられる.

3. 実験

3.1 作業課題

机の上に無作為に置かれた物体を HIRO のハンドカメラによって色画像認識し, その結果に基づき, 予めこちらが決めた

順番に 5 色を正しく並べるという知識を獲得するというものであり, その状態からの指定した順番と実際に取得した物体の順番の逸脱度をペナルティとした報酬を与える.

3.2 画像処理に基づく物体取得

作業課題に従って, 机の上に無造作に置いてある物体を選択された色に合わせてランダムもしくは greedy 選択によって探し出し, 物体を把持した後, 指定した場所に物体を移動させ, 併せて, 獲得した物体の色を記録する. 図 3 に画像処理に基づく物体取得の様子を示す.

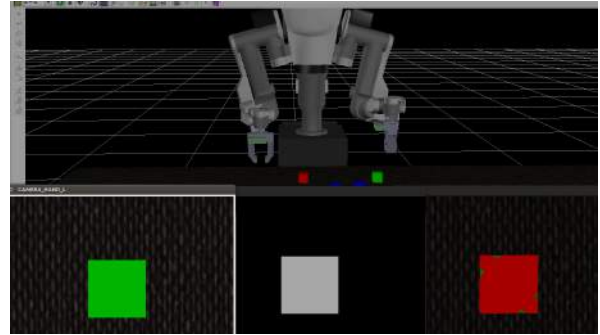


図 3: 画像処理に基づく物体取得の様子

3.3 Q 学習による行動知識獲得

状態および行動の表現形式は共に同形であり, (取得順番, 色) として与えられる. また, 現在の状態は, 過去に取得した物体の情報を保有しているとするため, 同じ色の選択は行われない. エージェントの行動選択方法として, ϵ -greedy 選択を用いた. また, 今回並べる色の決定方法について, Q 学習のシミュレーションによってのみ決定する方法と実世界とのインタラクションを組み込んだ Q 学習によって決定する方法の 2 種類の探索方法が考えられるため, 以下に詳しく示す.

3.3.1 シミュレーションによる色の決定

画像処理による物体の色認識を行わず, シミュレーションによる Q 学習によって取得する物体の色の順番を決定する. あらかじめ何色の物体が選択候補であるか指定しておき, その中から探索する色を考慮する. この際, 取得する物体の色は Q 学習によってランダムもしくは greedy 選択によって探し出す. この方法では, ロボットのハンドカメラにより取得された物体の色を判定した後に物体を把持しに行くという行為を起こさず, Q 学習によって決められた色の物体を把持しに行く.

3.3.2 実世界とのインタラクションによる色の決定

何色の物体があるか事前にロボットに教示せず, ハンドカメラに映し出された画像から何色の物体が机の上にあるのかを認識し, 探索する色を考慮する. この方法では実世界の情報に基づいて何色の物体があるのかを認識し, 取得する順番を決める. 画像情報から判定された色の物体を把持しに行くことから, 実世界のフィードバックなしには行うことができない. また, シミュレーションによる色の決定に比べて, 各色全て, 毎回カメラに何色が映っているかを確認しているため, 時間を要する. 実際に青色の物体, 緑色の物体の色を抽出している様子を図 4, 5 に示す. 手の移動による物体の探索と, ハンドカメラによる物体の色認識の作業を繰り返す事により, 物体の色情報の取得を行う.

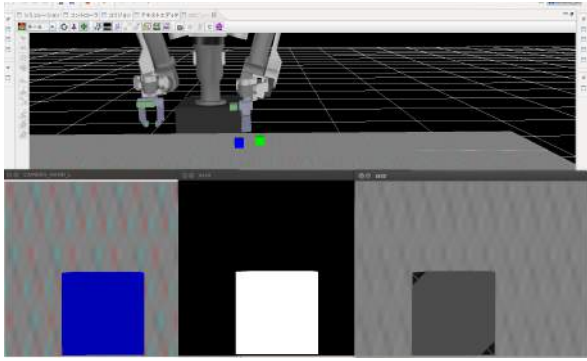


図 4: 青色の二値化処理により青色の物体を認識

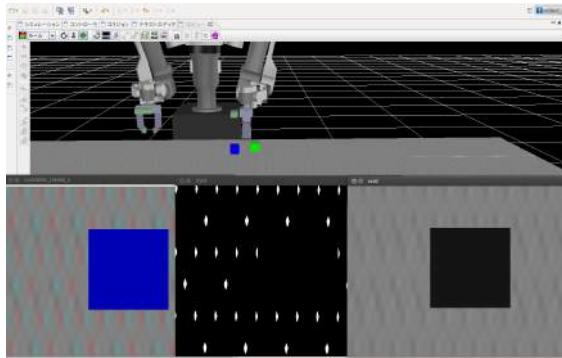


図 5: 青色の二値化処理により緑色の物体は認識されない

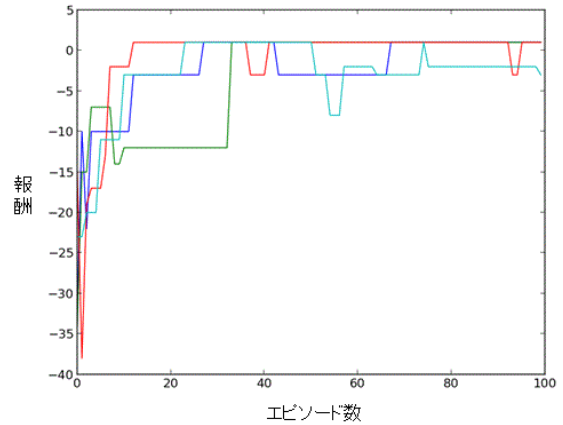


図 6: エピソード回数における報酬の遷移状況

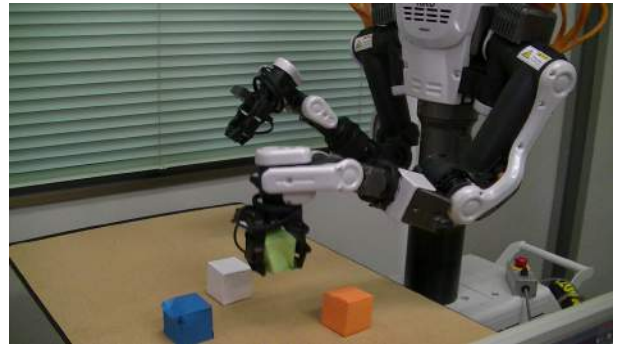


図 7: 実機での実験の様子

3.4 シミュレータを用いた実験

Q 学習においてエピソード回数に対する報酬の変化を見ることにより、学習の収束状態を確認した (図 6)。図 6 より、約 20 回ほど学習させた辺りから異なる結果が出ることもあるが、正解が現れ始め、およそ収束し始めることがわかる。各学習において、収束にバラつきが生じているが、これは作業課題が単純であり、状態の価値の変動が大きいことによって生じると考えられる。

画像処理に基づき色付きの物体を希望する順番で取得する行動知識を、Q 学習を通じて獲得できることをシミュレータ上で確認した。また、今回の実験では Q 学習により色を算出する方法で色決定を行っている。

3.5 実験結果と考察

シミュレータ上で確認した Q 学習を実際に実機に適用したところ、指定した色を発見、把持し、物体を探索する範囲の外に移動させることができた (図 7)。また、今回は物体の色判定には、3.3.1 のシミュレーションに基づいて行う方法を採用した。強化学習アルゴリズムとして、Q 学習を採用したため報酬の与え方が逐次的であり、正解を導き出すのが速かったと考えられる。

実験の中で探索する物体の色の決定方法として、Q 学習による行動知識獲得方法と画像情報によって色の抽出をする方法の 2 種類の検討を行った。前者に対し後者では、Q 学習にシミュレーションのみで取得する物体の色を決めるのとは異なり、実世界とのインタラクションを踏まえて色を決めているため処理に時間がかかることがわかった。

4. おわりに

本研究においては、画像認識に基づくヒューマノイドロボット HIRO の行動知識獲得を強化学習を用いて行うための基礎的な実験を行った。具体的には、HIRO に備え付けられたカメラから得られた画像中に映る物体の領域および色の認識を行い、物体を把持し、Q 学習のよって予め決められた順番で物体を取得することに成功した。今後の課題として、より現実世界の実験的な問題に対応するために、画像認識のバリエーションを増やすと共に逐次的な画像認識に基づく行動知識の獲得手法を開発していくとともに強化学習を用いたより複雑な課題についても検討していきたい。

謝辞

本研究は、JSPS 科研費 26280096 の助成を受けて実施した。

参考文献

- [1] Watkins, C.J.C.H., Learning from Delayed Rewards. PhD thesis, Cambridge University, Cambridge, England. 1989.
- [2] 浅田稔, 野田彰一, 依積田健, 細田耕, 視覚に基づく強化学習によるロボットの行動獲得, 日本ロボット学会誌, Vol.13, No.1, pp.68-74, 1995.
- [3] 浅田稔, 野田彰一, 細田耕, ロボットの行動獲得のための状態空間の自律的構成, 日本ロボット学会誌, Vol.15, No.6, pp.886-892, 1997.