6R-08

画像認識に基づくロボットの行動を制御する強化学習の取組み

恒川英里 † 小林一郎 † 麻生英樹 ‡ 長井隆行 § 中村友昭 § 持橋大地 ¶
† お茶の水女子大学 ‡ 産業技術総合研究所 § 電気通信大学 ¶ 統計数理研究所

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存在す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て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．具体例として，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 作業課題

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 HIRO を用いる．HIRO の腕に取り付けられたハンドカメラを用いて，テーブル上に置かれている色付きの物体の画像を取得する．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．HIRO はテーブル上の物体を取って来た際に，正解となる順番と比較し相当する報酬を獲得し，最終的に正解の順番にとってくるという行動知識を獲得する．図 1 に作業課題の概観を示す．

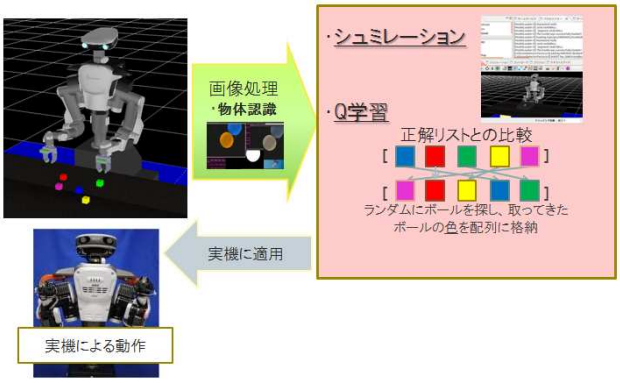

図 1: 作業課題の概観

An Approach to Robot Control based on Image Recognition using Reinforcement Learning
Eri TSUNEKAWA†, Ichiro KOBAYASHI†, Hideki Asoh‡
Takayuki Nagai§, Tomoaki Nakamura§, Daichi Mochihashi¶

2.2 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，二値化処理および輪郭抽出処理を行う．それにより物体の座標推定を行い(図 2)，その座標に手を移動し物体の把持を行う．

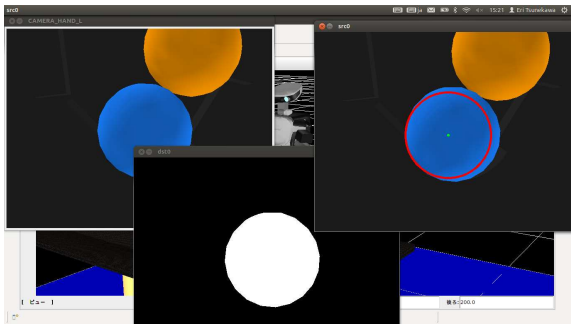

図 2: HIRO による画像処理

2.3 強化学習への定式化

2.3.1 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] 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 $E\{R_t | s_t = s, a_t = a\}$ で表現される行動価値を表し，α は学習率，γ は割引率を表す．

2.3.2 状態

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，以下に示す 3 つとする．

- 画像中に物体が映っている
- 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

2.3.3 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りだされた場合（ 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とになる．このことから本課題において対象となる動作は以下の4つとなる．

- ランダムに手を動かす
- 物体が映っている方向に手を動かす
- 物体を掴む
- 物体を掴まない

2.3.4 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あらかじ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列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 [赤, 青, 黄, 緑] 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，価値を赤: 4, 青: 3, 黄: 2, 緑: 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規順番との差分をペナルティして掛け，報酬とする．例えば，[緑, 黄, 赤, 青] という順番を取得したときその都度，報酬は 1 × (-3)，2 × (-1)，4 × (-2)，3 × (-2) として与えられる．

3 試作プログラム

3.1 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指定された色の順番に探し，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プログラムを作成した．図3に画像処理に基づく物体取得の様子を示す．

3.2 Q学習による行動知識獲得

作業課題を達成するQ学習の簡易なプログラムを試作し，課題に対する検討を行った．作業課題は，予め決まった順番に5色を並べるという知識を獲得するというものであり，その状態からの逸脱度をペナルティとした報酬を与える．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引き継いでいるとする．エージェントの行動選択方法として，ϵ-greedy 選択を用いた．

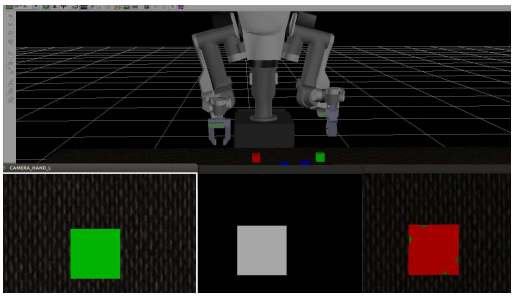

図 3: 画像処理に基づく物体取得の様子

3.3 実験結果と考察

画像処理に基づく物体取得プログラムについて，指定した色を見つけ，配列に収めることまではできるが，実際に取得を行う際のHIROが物体を掴もうとして手を移動する座標との乖離により失敗することがあった．物体を把持する際に物体とHIROの手の位置関係を正しく制御する必要があることがわかった．

Q学習プログラムにサンプルとしてランダムに取得した物体の色，順番の配列を与え，学習を行った結果，正解リストと同じ順序で物体を取得する知識を獲得することができた．しかし，実際に課題とする内容は画像に基づく物体の有無の判定およびそれに基づく行動において物体を探索するランダムな動作，物体の把持などがあるため，実際のHIROの行動知識の獲得にはこれらの要因を考慮する必要がある．

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の行動知識を強化学習を用いて行うための基礎的な検討を行った．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，物体を把持するプログラムの作成およびQ学習を用いて行動知識を獲得するプログラムを試作した．今後の課題として，試作プログラムをヒューマノイドロボットに適用し，実際にロボットを用いた知識獲得を行うと共に，現在片手を使って行っている作業を両手を使って作業できるようにするつもりである．

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，俵積田健，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.