

# 強化学習によるゴール守備行動の獲得

## Acquisition of a goal keeping behavior of a soccer robot by reinforcement learning

加藤 龍憲 (阪大) 正 鈴木 昭二 (阪大) 正 浅田 稔 (阪大)

Tatsunori KATO, Osaka University, 2-1, Yamadaoka, Suita, Osaka

Sho'ji SUZUKI, Osaka University

Minoru ASADA, Osaka University

**Abstract:** The authors apply Q-learning, one of major method of reinforcement learning, for a mobile robot to acquire a goal keeping behavior in the context of RoboCup. We equip an omnidirectional vision which consists of a hyperbolic mirror and a camera on the robot so that the robot can see its own goal and the ball in any direction. We also implement an attention control on the omnidirectional vision by controlling the focal length of the camera to track a moving ball. The robot with the omnidirectional vision and the attention control can learn a goal keeping behavior in short steps, the result is shown by simulation.

**Keywords:** behavior acquisition, reinforcement learning, omnidirectional vision, attention control, keeper

### 1 はじめに

ロボカップは、動的な環境化で作業するロボットチームの実現に必要な技術をサッカーを通じて追求することを目的としている<sup>1)</sup>。ロボカップにおける重要な研究課題の一つに、ロボットに実環境内で柔軟に対処できる行動を実現することが挙げられる。このような行動を獲得する手法として、ロボット自身が経験を通じて合目的な行動を獲得することのできる強化学習が注目されている<sup>2)</sup>。強化学習の枠組では、ロボットは観測した環境情報に基づいて自身のとるべき行為を決定する。この時、与えられた行動目的の達成度合に応じて環境からロボットに対し報酬が与えられる。これら一連の処理の繰り返しを通じて、ロボットは序々に得られる報酬を最大化するような行為を選択するようになり、最終的に合目的な行動を獲得する。

著者らは、視覚を持った単体の移動ロボットに強化学習の代表的な手法であるQ学習を適用し、ボールをゴールにシュートする行動を獲得させた<sup>3)</sup>。獲得された行動は実際のロボットに移植され、第1回ロボットワールドカップサッカー大会においてテストされた<sup>4)</sup>。しかしながら、ロボットに搭載されたカメラの視野が狭いため、転がるボールを見失いやすくシュートの成功率は高くはなかった。また、カメラで見える範囲が環境のごく一部であるためにロボットの得ることの情報が少なくシュート以外の行動の獲得は困難であった。

そこで、著者らはロボットの視覚に注視制御と全方位視覚を導入した上でQ学習を行いロボットの行動獲得を実現する。円錐や双曲面の鏡を利用した全方位視覚は、一度の観測で360度の情報を得ることができることから広く利用されており、全方位画像からのシーンの生成<sup>6)7)</sup>、移動ロボットのナビゲーション<sup>8)9)</sup>などに利用されている。これらの研究では、全方位画像から三次元のシーンや幾何モデルの再構成が中心であり、再構成をせずに画像情報を直接利用することは行われてこなかった。また、全方位画像における注視制御は取り扱われて来なかった。一般の視覚システムにおける注視制御は、カメラ本体を動かしたりパンやチルトを調整することにより実現されるが、全方位視覚の場合はカメラを動かすと物体の対応がとりに

くくなる。

本報告では、焦点距離の制御による全方位視覚のための注視制御を提案する。提案した手法を移動ロボットに組み込みQ学習により行動の獲得を行う。行動例としてサッカーにおけるゴール守備を取り上げ、シミュレーションにより提案した手法の有効性を示す。

### 2 タスクとロボットの設定

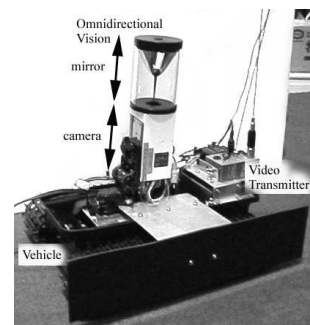


Fig.1 The robot

図1に著者らの移動ロボットを示す。ロボット本体は、PWS(Power Wheeled Steering)方式の移動機構を持ったラジコンを利用しており、これに全方位視覚と画像伝送器を搭載している。ロボットの大きさは、幅360[mm]長さ300[mm]高さ350[mm]であり、ロボットの最高速度は4.8[m/s]である。ロボットの制御はホストコンピュータから無線を通じて行われる。ロボットの行為としては図2に示すように、左右方向への直進、左右方向のその場回転、およびこれらの組合せによる8通りの行為と停止行為が予め実装されており、ホストコンピュータの遠隔操作により実行される。実行すべき行為はロボットから送られてくる画像情報を基にホストコンピュータ上で選択される。

全方位視覚は、双曲面鏡とCCDカメラにより構成され、双曲面鏡はその中心軸がCCDカメラの光軸と一致するよ

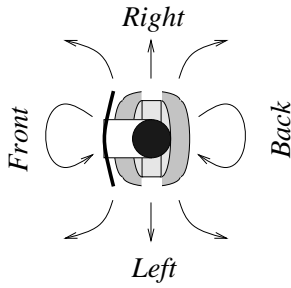


Fig.2 Actions of the robot

うに取り付けられている．ロボットの周囲にある物体は図3(a)に示すように双曲面鏡を通じてカメラの撮像面上に投影される．図3(b)に全方位視覚により得た画像の例を示す．全方位視覚は，ロボットが回転しても視野の中心が変化しないように，視野の中心がロボットの中心に一致するように取り付けられている．

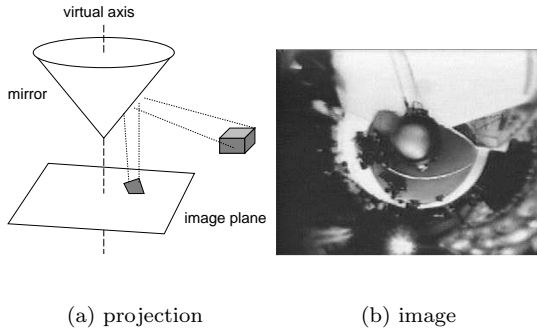


Fig.3 A sample of an omnidirectional vision

ロボットに与えるタスクは，サッカーにおけるゴール守備行動を扱い，これをQ学習によりロボットに獲得させる．ロボットは，全方位視覚によりゴールとボールを見ながら，ゴールめがけて転がるボールがゴールに入るのを防ぐ行動を獲得する(図4)．

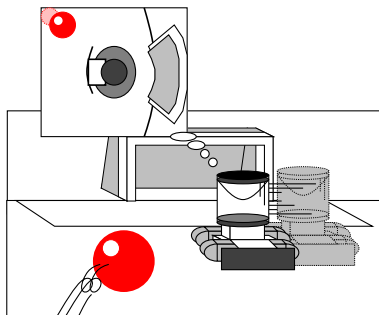


Fig.4 A goal keeping behavior

環境は，ロボカップの中型部門の競技場を基にし，競技場の半分の大きさである幅4575[mm]長さ4110[mm]の水平な平面を仮定する．ゴールの大きさは幅1500[mm]高さ

600[mm]とする．環境内にはゴールとボールのみが存在し，他のロボットは敵味方を含めて存在しないものとする．画像上ではゴールとボールは色により識別できるものとし，それぞれ青と赤に色分けされている．また，ボールの大きさは直径200[mm]で床面上を転がる．

### 3 全方位視覚における注視制御の実現

全方位視覚および観測する物体は水平面上に存在すると仮定し，座標系および各種パラメータを図5のように定義する．この時，実世界内の点 $P(R, \theta, Z)$ は撮像面上の点 $p(r, \theta)$ に投影され，それらには以下の関係が成り立つ．

$$Z = R \tan \alpha + c + h$$

$$\tan \gamma = \frac{b^2 + c^2}{b^2 - c^2} \tan \alpha + \frac{2bc}{c^2 - b^2} \frac{1}{\cos \alpha} \quad (1)$$

$$r = \frac{f}{\tan \gamma}$$

ここで， $a, b$ は双曲面の式 $\frac{R^2}{a^2} - \frac{Z^2}{b^2} = -1$ を決めるパラメータで $c = \sqrt{a^2 + b^2}$ の関係がある．ここで用いた全方位視覚では， $\alpha^2 = 233.3[\text{mm}]$ ,  $\beta^2 = 1135.7[\text{mm}]$ ,  $h = 250[\text{mm}]$ である．

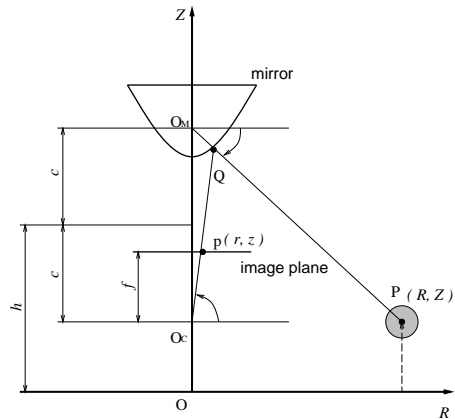


Fig.5 The projection of the hyperbolic mirror

著者らは全方位視覚上で注視を実現するために焦点距離の制御による注視制御を提案する．一般の視覚システムにおける注視は，注視する対象物を画像中心のように画像内の特定の位置に見続けるように，カメラのパンやチルトを制御することにより実現される．全方位視覚においては，対象物と画像の対応がとりにくくなるためにカメラを動かして注視を実現することは困難である．そこで，著者らは，焦点距離を変化させることにより注視する対象物を画像中心から一定の距離にある円周に見続ける方法を提案する(図6)．

画像中心からの距離 $r$ を一定とした時，物体までの距離 $R$ とカメラの焦点距離 $f$ の関係は式(1)より求められ，図7の通り図示される．この図より，物体までの距離が変化しても焦点距離を制御することにより画像中心からの距離を一定に保つことが可能であることが示される．そこで，注視制御を以下の式に従って行う．

$$u_f = K(I r_d - I r) \quad (2)$$

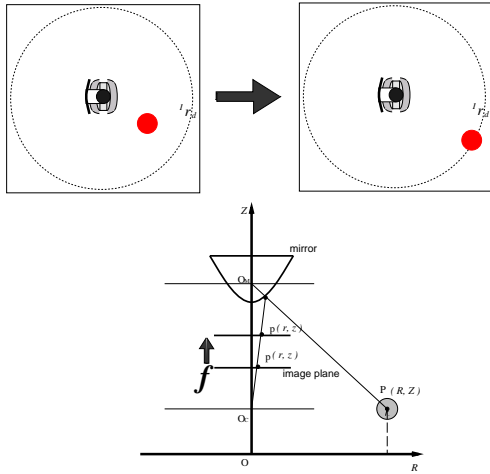


Fig.6 Attention control on the omnidirectional vision

ここで、 $u_f$  はカメラの焦点距離の入力値、 $I_{r_d}$  は注視物体の画像上での目標値 (注視円半径)、 $I_r$  は現在の画像上の位置である。例えば、図6に示すように、 $I_r$  が小さい場合には  $f$  を大きくすることにより  $I_{r_d}$  に近づけることができる。

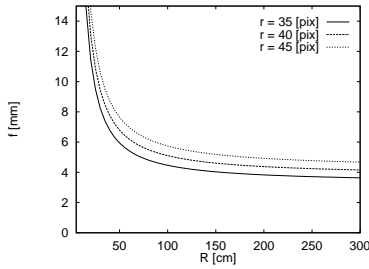


Fig.7 The relation between distances and the focal length

#### 4 学習による行動の獲得

Q学習を用いてロボットに行動を獲得させるためには、ロボットの得られる視覚情報から行動に対して有効な特徴に着目し、状態を定義しなければならない<sup>3)</sup>。ゴール守備行動の場合は、ゴールとボールの画像上の位置、およびボールの移動方向が重要である。表1に定義した状態を示す。

注視制御を行わない場合は、ボールに関しては画像中心からの距離と画像上の方位に関する状態が必要である。また、ボールの移動方向を知るために、距離と方位のそれぞれに関する変化を状態としなければならない。したがって、ボールの画像上での方位を図8(a)に示す8状態とし、方位の変化を3状態(時計回りに変化した, 変化無し, 反時計回りに変化した)と定義した。ボールの距離は8(b)に示す3状態(遠い, 中間, 近い)とし、距離の変化は3状態(遠ざかる, 変化なし, 近づく)と定義した。ゴールに関しては方位をボール同様図8(a)に示す8状態とし、距離を2状態(遠い, 近い)と定義した。その結果, 注視制御を行わない場合の総状態数は、 $8 \times 3 \times 3 \times 3 \times 8 \times 2 = 3456$  となった。

注視制御を行う場合は、注視制御により画像中心からのボールの距離は一定となるため、状態を定義する必要がなくなる。そのため、ボールに関しては画像上での方位と方位の変化に対してのみ状態を定義すればよく、学習のための状態を削減することができる。ゴールに関しては注視制御なしの場合と同様に距離と方位が必要である。ただし、ゴールの見え方は注視制御の有無により変わる。注視制御を行わない場合には画像上の距離はゴールとロボットとの距離に対応するが、注視制御を行う場合はボールとゴールとの相対距離に対応する。結果的に注視を行う場合の総状態数は、 $8 \times 3 \times 2 \times 8 = 320$  となった。

	注視制御なし	注視制御あり
画像上でのボールの方位	8	8
ボールの方位の変化	3	3
画像中心からのボール距離	3	-
ボールの距離の変化	3	-
画像上でのゴールの方位	8	8
画像中心からのゴール距離	2	2
総状態数	3456	320

Table 1 Substates

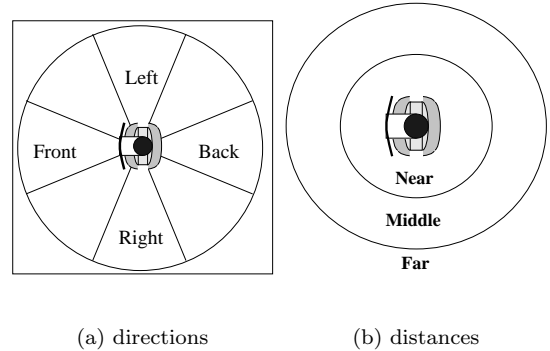


Fig.8 Substates

#### 5 実験結果

シミュレーションにより、ロボットにゴール守備行動を学習させた。ロボットとボールの初期値は図9に示す通り設定した。ロボットの初期値はゴール前のある空間、ボールの初期値はゴール中心とコーナを半径とする円周上に配置した。ボールはゴール目がけてロボットの最高速度の半分の速度で減衰することなく転がる。一試行は、ボールがゴールに入った場合とボールがフィールド外に出た場合に終了する。報酬はボールがゴールに入った場合には負の報酬を与え、ゴール守備をする上で最悪の行動を選択しないように学習を行わせた。

注視制御を行う場合と行わない場合のシミュレーション結果を図10に示す。図は、10000試行, 20000試行, 30000試行, 4000試行後の学習結果によるシュートの阻止率を示す。これより、注視点制御を行いながら学習した場合の方がより少ない試行で行動が獲得できることが示された。

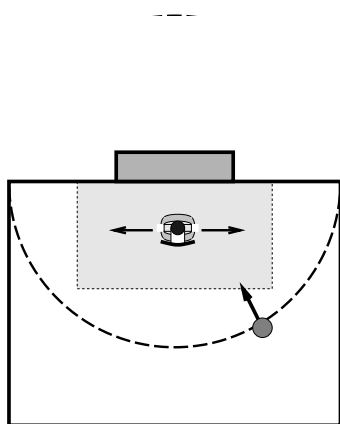


Fig.9 Initial position

シミュレーションで獲得した行動を実ロボットに適用した結果、ロボットは図11(a)-(f)に示すように行動した。

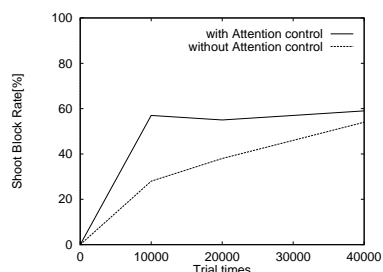


Fig.10 Result

## 6 まとめ

本稿では、全方位視覚に焦点距離の制御による注視制御を実現し、これを搭載した移動ロボットにサッカーのゴール守備行動を学習させた。注視制御を行った場合は、行わない場合に比べてより短時間で学習することができた。このことは、注視制御が状態空間の削減に貢献し学習の時間を短縮できる場合があることを示している。しかしながら、今回の結果は一つのケーススタディであり、獲得すべき行動とそれに必要な状態空間および状態空間を縮小するために適した注視制御の関係は明らかにされていない。今後、この点を明らかにしていくことが課題である。

また、今回は他のロボットの存在しない簡単な環境下で学習を行ったが、今後より多くのロボットの存在する複雑な環境への適応も重要な課題である。

## 参考文献

- [1] H. Kitano, M. Asada, Y. Kuniyoshi, I. Noda, E. Osawa, and H. Matsubara. "robocup: A challenge problem of ai". *AI Magazine*, 18:73-85, 1997.
- [2] Connel, J. H., Mahadevan, S.: *Robot Learning*. Kluwer Academic Publishers (1993)
- [3] 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

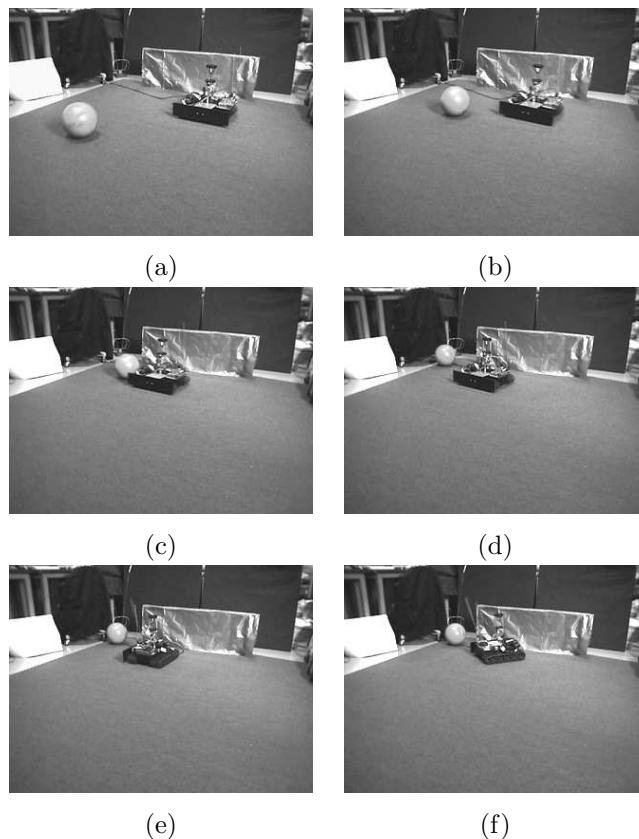


Fig.11 A sequence of behavior

ト学会誌, Vol.13, No.1, pp.68-74,1995.

- [4] 鈴木昭二・浅田稔: "学習によるロボットの行動獲得 - サッカーで試みるロボカップへの挑戦 -", 電気学会誌, Vol.118, No.1, pp.21-24, 1998.
- [5] Y. Yagi and S. Kawato. Panoramic scene analysis with conic projection. In *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems 1990 (IROS'90)*, 1990.
- [6] S. Nayar. Omnidirectional video camera. In *Proc. of 1997 Image Understanding Workshop*, pages 235-241, 1997.
- [7] V. N. Peri and S. Nayar. Generation of perspective and panoramic video from omnidirectional video. In *Proc. of 1997 Image Understanding Workshop*, pages 243-245, 1997.
- [8] K. Yamazawa, Y. Yagi, and M. Yachida. Obstacle avoidance with omnidirectional image sensor hyperomni vision. In *Proc. of IEEE Int. Conf. on Robotics and Automation*, pages 1062-1067, 1995.
- [9] H. Ishiguro. Distributed vision systems: A perceptual information infrastructure for robot navigation. In *Proc. of IJCAI-97*, pages 36-41, 1997.