

強化学習によるクルーズ型 AUV の ピッチ角選択アルゴリズム ~高速かつ低高度な海底追従を目指して~

Pitch Control Method for a Cruising-AUV by Reinforcement Learning ~For High Speed and Low Altitude Terrain Following~

野口 侑要, 巻 俊宏

Yukiyasu Noguchi, Toshihiro Maki

東京大学生産技術研究所

Institute of Industrial Science, the University of Tokyo

Abstract: Low altitude survey of seafloor is important in various fields, such as biological research or resource survey. An AUV (Autonomous Underwater Vehicle) plays an important role. In this paper, we propose a new terrain following method, in which a reinforcement learning agent sets an appropriate pitch reference. We expect the new method results in self-adaptive to the environment. We conducted a simulation based analysis, in which the AUV traveled at a high surge speed (~2 m/s) in various sonar echo level environment.

1 はじめに

海底には貴重な鉱物資源や生物資源が存在し、今後の人類の発展のために役立てられることが期待されている。それらの資源調査のため、カメラなどの光学機器を用いた観測が求められているが、海中においては光の減衰が激しく、対象物と数メートルの距離に近づかなければならないと言われている。人間が直接潜ることのできる深さは限られているため、さまざまな無人海中ロボット[1]が運用されており、近年は遠隔操縦ロボット (ROV) に加えて、無索で運用することができる自律型水中ロボット (AUV) も活躍している[2-4]。その中でも複数台のスラスタを持ち、細かな姿勢制御をすることが可能なホバリング型 AUV が海底画像観測に使われつつある。起伏が激しい海底において、細かな姿勢制御をすることが可能なホバリング型 AUV は、海底に衝突するリスクが低く、安定して運用することができる。一方でホバリング型 AUV は、海底構造物との衝突を

避けるため、毎秒数 10cm 程度の低速度で運用されるのが一般的で、調査可能領域や調査必要時間といった効率の点で問題であった。

本研究では、AUV の制御に強化学習[5-6]を導入することで、海洋の環境変化に対応しながら、低高度かつ高速な海底追従を行う手法の開発を行っている。強化学習とは環境に対する試行錯誤的な働きかけと、それに対して環境から受ける情報を通じて学習していく機械学習の一種である。強化学習は、学習を行う主体であるエージェントと、エージェントが働く環境という、二つの枠組みで議論されることが多い。エージェントは、環境から現在おかれている状況を伝えられ、その中で行動することによって、環境に対して働きかけをする。その行動の結果得られた次の状態と報酬によって、エージェントはそれぞれの状況において、これから得られる報酬を最大化するための行動指針を学習する。AUV の制御に強化学習を用いることにより、次の2点の貢献ができると考えられる。

・事前に知識のなかった海底環境、AUV の状況への適応性の向上

・さまざまな海底環境を模擬したシミュレータ内においてエージェントを学習させることにより、人間では思いつかなかった新発想の制御方針の獲得。

本稿においては、AUV が環境に適応するという効果に焦点をあて、低高度高速航行における AUV のピッチ角制御手法を提案する。AUV による低高度高速航行を実現させる際の問題点として、環境情報が乏しい点が挙げられる。高速で AUV が海底追従を行う際には、ある程度前方の環境を前もって観測しなければならないが、水中におけるほぼ唯一の遠距離探知手段である音響は、環境によってエコーレベルが異なり、どのような環境でも確実に障害物などの環境情報を得ることは、困難である。また、例えば同じ高度 2 m から観測するミッションでも、その高度の重要性はその目的によって異なる。例えば、海底に衝突してでも高度 2m での画像が必要なケースもあれば、衝突回避を最優先すべきミッションも考えられる。強化学習の採用により、そのようにミッション毎に変わる環境と目的のもと、適した制御を AUV が自ら選択することができるかと期待される。

2 提案手法

本手法において AUV は障害物回避をピッチ角制御によってのみ行うものとする。また、AUV はクルーズ型と呼ばれる高速航行に適した形式を想定する。また、前方の環境観測のためには、ソナーの中でも安価であり、広く使われているスキャニングソナーを使うことを想定する。また何らかの手法で海底からの高度を測定できるものとする。

ピッチ角選択アルゴリズムは、パラメータ調節を強化学習手法の一つである Sarsa(λ)[7]で行う。Sarsa(λ)のアルゴリズムを Fig. 1 に示す。変数の定義は以下の通りである。

- ・ s , state
- ・ a , action
- ・ δ , TD (Temporal difference) error
- ・ r , reward
- ・ γ , discount rate
- ・ α , step-size parameter
- ・ $Q(s,a)$, action-value function
- ・ $e(s,a)$, eligibility trace
- ・ λ , trace-decay parameter

```

Initialize  $Q(s', a')$ 
For each episode:
  Initialize  $e(s, a)$ ,  $s$ ,  $a$ 
  For each time step:
    Perform  $a$ , and observe  $r$ ,  $s'$ 
    Choice  $a'$ 
     $\delta \leftarrow r + \gamma Q(s', a') - Q(s, a)$ 
     $e(s, a) \leftarrow e(s, a) + 1$ 
    For all  $s, a$ :
       $Q(s, a) \leftarrow Q(s, a) + \alpha \delta e(s, a)$ 
       $e(s, a) \leftarrow \gamma \lambda e(s, a)$ 
     $s \leftarrow s'$ ;  $a \leftarrow a'$ 
  
```

Fig. 1. Sarsa(λ) アルゴリズム

AUV はソナーからある閾値 D 以上のエコーレベルが得られた地点を海底と認識する。そしてその時点で認識している海底から理想とする高度 H 分高い地点をすべて回避できる最小のピッチ角を制御指令値として出力する。エコーレベルは対象物の材質や形状、入射角に依存する。さらに、音響ビームには幅があるため、実際には構造物が存在しない地点からも高いエコーレベルが得られることがあり、画一的に閾値 D を設定することは困難である。そこで、Sarsa(λ)エージェントにより最適な閾値 D を推定する。

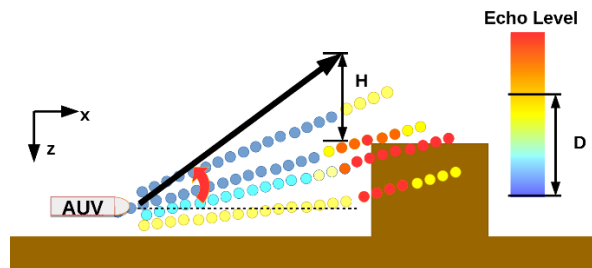


Fig. 2. ピッチ角選択アルゴリズム

状態 s としては、その時点での、高度とピッチ角を与える。

報酬 r は、AUV が指定高度で海底を観測し、かつ海底に衝突しないように、以下の Fig. 3 のように定義する。

```

If AUV crashes obstacles:
   $r \leftarrow$  (Crash penalty)
If altitude > (Max limit altitude to take photo):
   $r \leftarrow$  (No photo penalty)
Else:
   $r \leftarrow$  (Photo value)
  
```

Fig. 3. 報酬 r の定義

Crash penalty や No photo penalty, Photo value という変数はミッションの特性に応じて、人間が設定するパラメータである。

3 シミュレーション

提案手法の有効性を確認するため、本手法により閾値 D の調節を行う AUV と行わない AUV の二つをシミュレーション環境で比較する。シミュレーションの途中で海底のエコーレベルを約半分に低下させ、それぞれの AUV の振る舞いにどのように影響を与えるのか調べる。シミュレーション環境には、ROS (Robot Operation System) [8] と互換性の高い Gazebo シミュレータ [9] を用いる。当研究室の AUV のハードウェアインタフェースを担当する ROS ノードと互換性のあるノードを Gazebo シミュレータの中に構築することによって、本提案手法部分を、実機と変わらない形で自由に動作させることができる。

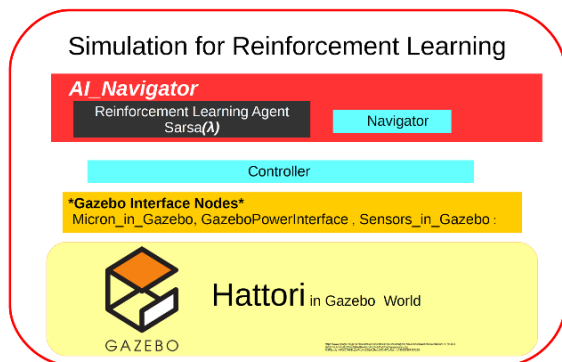


Fig. 4. AUVシミュレータの構成

シミュレーションに用いる機体のモデルには、当研究室で開発中の AUV Hattori を用いる。Hattori は流体抵抗の小さい機体形状と強力なスラストにより、約 2 m/s で運用することができる。前方にはスキャニングソナー、下方には観測用のカメラが搭載されている。



Fig. 5. AUV Hattori

Table 1, Hattori の仕様

Size	1.02 m (L) x 0.2 m (H) x 0.48 m (W)
Mass	16 kg
Speed	~ 2.0 m/s

シミュレーション環境において、AUV は約 200m の海底を模擬したフィールドを高速で (~2 m/s) 直進する。スキャニングソナーに関しては、Gazebo シミュレータのレーザプラグインを重ねることで厚みのある音波とし、レーザプラグインから得られる距離情報から、エコーレベルへ変換している。

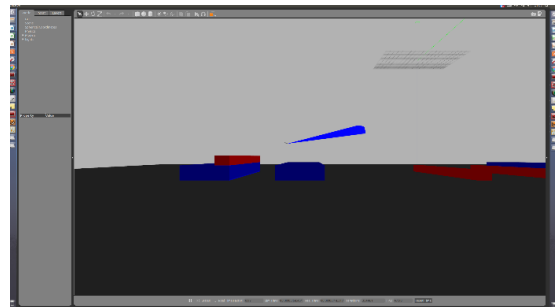


Fig. 6. AUVシミュレータのGUI

フィールドには障害物として、直方体がランダムに設置されている。AUV は最大で約 7m の高さを持つ障害物をのりこえなければならない設定である。フィールドを終端まで進み終わったら、AUV は開始位置までもどされ、障害物は新たにランダムに配置される。開始地点から終端までの航行を 1 エピソードと数える。前述した環境の変化は、0 エピソードから 1 エピソードに切り替わるタイミングで行う。本提案手法を用いたシミュレーションを Case 1, 強化学習を用いないバージョンのシミュレーションを Case 2 と呼ぶ。

Sarsa(λ) エージェントは 5Hz で動作させる。また、報酬 r に関するパラメータを以下のように与える。

- Max altitude to take a seafloor photo = 4.0 [m]
- Crash penalty = -100
- No photo penalty = - 0.02

海底画像は 2m 高度から撮るのが望ましいという想定で、H=2m とし、Photo value を以下のように定義している。

$$(\text{Photo value}) = 0.1 * (2.0 - \text{absolute}(2.0 - \text{altitude})) \quad (1)$$

3 結果

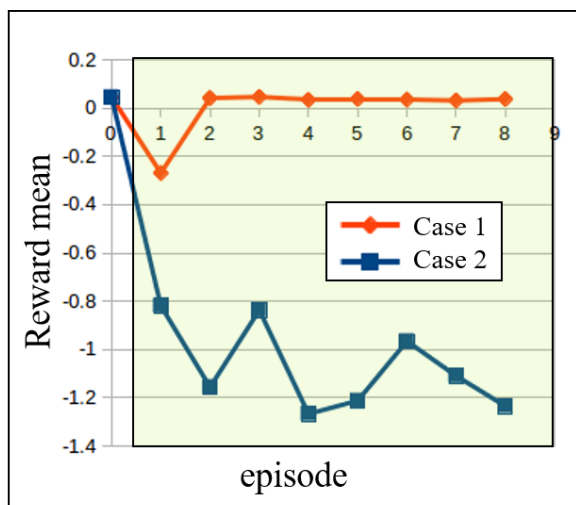


Fig. 7. 平均報酬

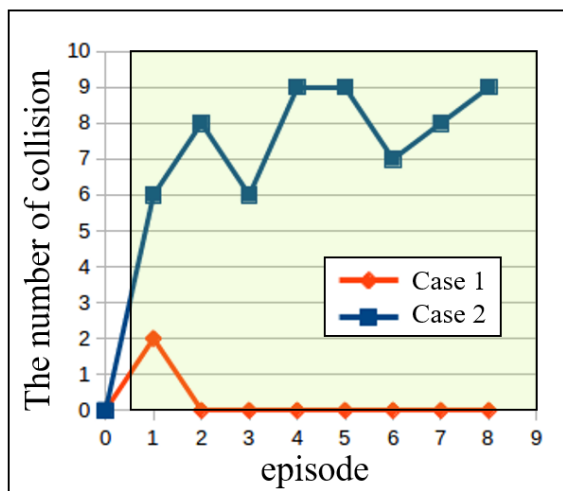


Fig. 8. 衝突回数

Fig.7 は各エピソードにおける、1動作周期あたりの平均報酬を表し、Fig.8 は各エピソードにおける、総衝突回数を表している。エコーレベルが下がる前は、海底に衝突することなく、高い報酬が得られている。しかし、エコーレベルが下がったあと、Case 2 においては海底へ衝突することが多くなり、平均報酬が大きく下がっている。一方で、Case 1 では、エコーレベルが変化し直後は海底に衝突して報酬が下がるものの、その後すぐに回復している。

これらの結果から提案手法により、環境の変化に対応し、新たに適切な閾値を設定することに成功したと言える。

4 おわりに

本稿では強化学習によりソーナーの海底検出閾値をリアルタイムに学習することで、海洋の環境変化に対応しながら、AUVにより低高度かつ高速な海底追従を実現できることを示した。本稿で扱った状態空間や行動空間は小さいものであるが、将来的に、シミュレータによる事前学習や、ディープラーニングなどといった関数近似手法を用いれば、強化学習の利点をより複雑な AUV の制御問題に適応できると考えられる。

参考文献

- [1] 浦 環, 高川 真一: 海中ロボット, 成山堂書店, (1997)
- [2] T. Maki, A. Kume and T. Ura, "Volumetric mapping of tubeworm colonies in Kagoshima Bay through autonomous robotics surveys," *Deep-Sea Research I*, 58, pp. 757-767, (2011)
- [3] Y. Nishida, K. Nagahashi, T. Sato, A. Bodenmann, B. Thornton, A. Asada, T. Ura, "Autonomous Underwater Vehicle "BOSS-A" for Acoustic and Visual Survey of Manganese Crusts," *Journal of Robotics and Mechatronics*, 28, 91-94, (2016)
- [4] S. McPhail, M. Furlong and M. Pebody, "Low-altitude terrain following and collision avoidance in a flight-class autonomous underwater vehicle," *Proceedings of the Institution of Mechanical Engineers, Vol. 224, Part M: Journal of Engineering for the Maritime Environment*, pp. 279-292, (2010)
- [5] 三上 貞芳, 皆川 雅章: 強化学習, 森川出版株式会社, (2011)
- [6] 牧野 貴樹, 澁谷 長史, 白川 真一: これからの強化学習, 森川出版株式会社, (2016)
- [7] Rummery, G. A., and Niranjan, M. "On-line Q-learning using connectionist systems," *Technical Report CUED/F-INFENG/TR 166*, (1994)
- [8] ROS, <http://wiki.ros.org/>
- [9] Gazebo シミュレータ, <http://gazebosim.org/>