Robotics and Autonomous Systems 107 (2018) 71–86

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot

Adaptive low-level control of autonomous underwater vehicles using deep reinforcement learning

Ignacio Carlucho [a,b], Mariano De Paula [a,*], Sen Wang [b], Yvan Petillot [b], Gerardo G. Acosta [a]

[a] INTELYMEC group, Centro de Investigaciones en Física e Ingeniería del Centro CIFICEN –UNICEN –CICpBA –CONICET, Argentina
[b] School of Engineering & Physical Sciences Heriot-Watt University, EH14 4AS, Edinburgh, UK

HIGHLIGHTS

- Adaptive low-level control strategy of autonomous underwater vehicle.
- Deep reinforcement learning to solve a continuous control problem.
- Real experiments demonstrated the feasibility of deep RL for AUV low-level control.
- Only raw sensory information is used for the deep RL actor-critic architecture.

ARTICLE INFO

Article history:
Received 22 February 2018
Received in revised form 2 May 2018
Accepted 30 May 2018
Available online 15 June 2018

Keywords:
Autonomous robot
Deep reinforcement learning
AUV
Adaptive low-level control

ABSTRACT

Low-level control of autonomous underwater vehicles (AUVs) has been extensively addressed by classical control techniques. However, the variable operating conditions and hostile environments faced by AUVs have driven researchers towards the formulation of adaptive control approaches. The reinforcement learning (RL) paradigm is a powerful framework which has been applied in different formulations of adaptive control strategies for AUVs. However, the limitations of RL approaches have lead towards the emergence of deep reinforcement learning which has become an attractive and promising framework for developing real adaptive control strategies to solve complex control problems for autonomous systems. However, most of the existing applications of deep RL use video images to train the decision making artificial agent but obtaining camera images only for an AUV control purpose could be costly in terms of energy consumption. Moreover, the rewards are not easily obtained directly from the video frames. In this work we develop a deep RL framework for adaptive control applications of AUVs based on an actor-critic goal-oriented deep RL architecture, which takes the available raw sensory information as input and as output the continuous control actions which are the low-level commands for the AUV's thrusters. Experiments on a real AUV demonstrate the applicability of the stated deep RL approach for an autonomous robot control problem.

1. Introduction

Autonomous underwater vehicles are revolutionizing the oceanic research with applications on a vast number of scientific fields such as marine geoscience, biology and archeology but also in the private sector such as the oil and gas industry [1,2]. Over the years, there have been intensive efforts toward the development of autonomous control strategies for AUVs [3]. Autonomy implies that an entity can act independently according to its own criterion and it is an essential feature for engineering systems in large and uncertain environments [4]. In this sense, adaptive low-level control techniques have arisen as a way to provide autonomy to AUVs allowing them to operate in hostile environments [5].

Classical control theory has evolved in a variety of methods for low-level AUV control. Several versions of the well-known PID controller have been developed and used for AUV control. To name a few, in the early work of Jalving [6] a simple proportional derivative controller was proposed for AUV steering control. Fjellstad and Fossen [7] designed a PID controller for position and attitude tracking of an AUV and the global convergence of their proposal was proven by Barbalat's lemma. More sophisticated proposals can be found in the work of Valenciaga, et al. [8] where a proportional integrative controller for multiple inputs and multiple outputs (PI-MIMO) was formulated to command the rudder and the propeller of an AUV. In the work of Sutarto and Budiyono [9] a linear

* Corresponding author.
E-mail addresses: ignacio.carlucho@fio.unicen.edu.ar (I. Carlucho), mariano.depaula@fio.unicen.edu.ar (M. De Paula), s.wang@hw.ac.uk (S. Wang), y.r.petillot@hw.ac.uk (Y. Petillot), ggacosta@fio.unicen.edu.ar (G.G. Acosta).

https://doi.org/10.1016/j.robot.2018.05.016
0921-8890/

parameter varying (LPV) control strategy based on linear fractional transformation to formulate a robust gain schedule strategy for robust longitudinal control of an AUV was developed. To deal with the AUV modeling uncertainties and the saturations of the control actions imposed by the AUV actuators, Sarhadi et al. [10], proposed an adaptive PID formulations with anti-windup compensators and then the stability was analyzed by Lyapunov theory and the proposed control technique was implemented in an onboard computer to be checked in a real-time dynamic simulation environment.

When model estimation accuracy could be imprecise and the system nonlinearities are considered, Lyapunov-based algorithms have many advantages for control formulations. An example can be found in Ferreira et al. [11] where several independent controllers have been developed, based only on Lyapunov theory, to perform decoupled motions of an AUV. In the work of Lapierre and Jouvencel [12] a nonlinear robust control formulation resorting to Lyapunov-based techniques was presented. In this case a virtual target principle was used to design an asymptotically convergent kinematic control, relying on a switching control strategy for the dynamic parameters. However, the disturbance rejection was not explicitly addressed in the formulation and the authors have explicitly recognized that further research is needed. In another way, developments coming from nonlinear control designs have been made where linear transformations were used to solve Linear Quadratic and Gaussian regulators (LQR and LQG, respectively) as in the work of Wadoo et al. [13] where a system linearization is carried out for the control of the kinematic model of an AUV and then a LQG was formulated as a H-2 optimization problem. Geranmher et al. [14] considered a general fully coupled AUV and applied nonlinear suboptimal control, where the state-dependent Riccati equation was used to generate a suboptimal path solution. In the work of Fischer et al. [15] a continuous robust integral of the sign of the error control was used to compensate for uncertain, nonautonomous disturbances for a coupled and fully-actuated underwater vehicle. Moreover, semiglobal asymptotic stability was proven by a Lyapunov-based stability analysis.

Underwater vehicle hydrodynamics are highly non-linear with uncertainties that are difficult to parameterize and, in addition, unknown disturbances are usually present as are typical of aquatic environments. For these reasons, researchers have resorted to adaptive controllers and have often included the dynamical model or have estimated the system parameters in the formulation of the controllers. Early, Fossen and Fjellstad [7] discussed the performance of the adaptive control laws for controlling underwater vehicles. Afterward, several adaptive PID formulations have been proposed as in works of Antonelli et al. [16] where different adaptive versions based on PID control laws were formulated with an adaptive compensation of the dynamics. However, in such proposals the control gains must be adjusted manually, first in simulation and then with the real system during its operation [17]. An adaptive on-line tuning method for a coupled two-loop proportional controller of four degrees-of-freedom for an autonomous underwater vehicle is presented in the work of Barbalata et al. [18] where the gains of each controller are determined on-line according to the error signals. Rout and Subudhi [19] developed an adaptive tuning method for a PID controller using an inverse optimal control technique based on a NARMAX model for the representations of the non-linear dynamics. Other adaptive feedback controller was proposed by Narasimhan and Singh [20] using LQR theory for the computation of the optimum feedback gain vector of the control system, in this case used for depth control of a low-speed underwater vehicle. These facts evidence a growing need for self-adapting controllers to environmental conditions.

To enhance the different control formulations researchers have turned their attention to artificial intelligence techniques to be incorporated in adaptive control formulations to develop real autonomous systems. Particularly, using artificial neural networks (ANNs) in AUV control formulations has the advantage that the dynamics of the AUVs do not need be fully known and ANNs can learn a full, or partial, model of the nonlinear dynamics which can in turn be used for the controller design [21]. In Shi et al. [22] a hybrid control approach for AUV depth control has been proposed using the Lyapunov theory approach for the synthesis of an adaptive controller and an ANN was employed to model the depth dynamics. A dual closed loop control system was proposed in [23] where a bio-inspired model for velocity control was used in an inner control loop and a sliding-mode controller was used in an outer tracking control loop which managed the position and orientation of an AUV. Also, a traditional Lyapunov stability analysis was carried out based on the AUV dynamic model. However, strong nonlinearities, as in underwater vehicles applications, make this analysis difficult. In this sense, after the development of the fuzzy logic many fuzzy control strategies were proposed for AUV control [24–27]. Briefly, fuzzy logic control makes a smooth approximation of a nonlinear system using a fuzzy inference system [28] consisting of a set of linguistic rules about the system behavior and membership functions which must be conveniently defined. In the work of Raeisy et al. [29] a simple fuzzy control formulation can be found with two fuzzy control loops, one that controlled the roll and yaw and the other the depth of the AUV, while incorporating an optimization procedure for the fuzzy parameters using the root mean square error between the input and the output as cost function. Recently, Khodayari et al. [30] have proposed a self-adaptive fuzzy PID controller for the attitude control of an AUV based on its previously obtained dynamic model from mechanical principles. Also, fuzzy control formulations for underwater vehicle-manipulator system (UVMS) were formulated in Esfahani et al. [31]. However, one disadvantage for using fuzzy control systems for AUVs is that subjective knowledge is required for the definition of the fuzzy rules and membership functions.

Other important branch with growing importance in the field of artificial intelligence for autonomous control systems is the RL paradigm [32]. Instead of supervised learning as ANNs, RL is a mixed approach between supervised and unsupervised learning using actor-critic approach with potential advantages for adaptive control formulations in robotics [33,34]. In a nutshell, RL algorithms are able to learn a control policy through the interactions between the system and its environment. RL algorithms can be formulated as model-free and/or model-based [35,36]. The former uses the experience from interaction to determine directly the optimal control policy [32,37] while the latter uses it to learn/update the current model of the system or to improve the value function and/or the policy directly [38].

Particularly, for AUVs relevant works have been developed using RL formulations. In the early work of Gaskett et al. [39] a model-free RL algorithm was developed to control the thrusters responses of an AUV. More recently, Carreras et al. [40] proposed a hybrid behavior-based scheme using RL for high-level control of an AUV. In this work a semi-online neural-Q-learning algorithm was formulated using a multilayer neural network to learn the internal continuous state–action mapping of each behavior. In the work of El-Fakdi et al. [41] an on-line direct policy search algorithm based on a stochastic gradient descent method with respect to the policy parameter space was proposed. In this formulation, the policy was represented by a neural network, where its weights were the policy parameters. The states of the systems were the inputs to the neural network and the outputs were the action selection probabilities [42]. Then, El-Fakdi and Carreras [43] developed a simulation-based actor-critic algorithm using policy gradient method to solve a cable tracking task. In this formulation an initial policy is learned off-line using a hydrodynamic model of the AUV.