

Proceeding Paper

Using Reinforcement Learning in the Path Planning of Swarms of UAVs for the Photographic Capture of Terrains [†]

Alejandro Puente-Castro ^{1,*} , Daniel Rivero ¹ , Alejandro Pazos ^{1,2}  and Enrique Fernandez-Blanco ¹ 

¹ Faculty of Computer Science, CITIC, University of A Coruña, 15071 A Coruña, Spain; daniel.rivero@udc.es (D.R.); alejandro.pazos@udc.es (A.P.); enrique.fernandez@udc.es (E.F.-B.)

² Biomedical Research Institute of A Coruña (INIBIC), University Hospital Complex of A Coruña (CHUAC), 15006 A Coruña, Spain

* Correspondence: a.puentec@udc.es

† Presented at the 4th XoveTIC Conference, A Coruña, Spain, 7–8 October 2021.

Abstract: The number of applications using unmanned aerial vehicles (UAVs) is increasing. The use of UAVs in swarms makes many operators see more advantages than the individual use of UAVs, thus reducing operational time and costs. The main objective of this work is to design a system that, using Reinforcement Learning (RL) and Artificial Neural Networks (ANNs) techniques, can obtain a good path for each UAV in the swarm and distribute the flight environment in such a way that the combination of the captured images is as simple as possible. To determine whether it is better to use a global ANN or multiple local ANNs, experiments have been done over the same map and with different numbers of UAVs at different altitudes. The results are measured based on the time taken to find a solution. The results show that the system works with any number of UAVs if the map is correctly partitioned. On the other hand, using local ANNs seems to be the option that can find solutions faster, ensuring better trajectories than using a single global network. There is no need to use additional map information other than the current state of the environment, like targets or distance maps.

Keywords: UAV swarm; path planning; reinforcement learning; Q-learning; artificial neural network; terrain



Citation: Puente-Castro, A.; Rivero, D.; Pazos, A.; Fernandez-Blanco, E. Using Reinforcement Learning in the Path Planning of Swarms of UAVs for the Photographic Capture of Terrains. *Eng. Proc.* **2021**, *7*, 32. <https://doi.org/10.3390/engproc2021007032>

Academic Editors: Joaquim de Moura, Marco A. González, Javier Pereira and Manuel G. Penedo

Published: 15 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

There are more and more applications for the collective use of Unmanned Aerial Vehicles (UAVs), more known UAV swarms. In addition to the advantages of the individually usage of these systems, the main motivation for swam usage is the reduction of flight time and operating costs together with increased fault tolerance [1]. Advances in the creation of algorithms [2] and telecommunications [3] allow us to have collective systems that are practically autonomous in their entirety. Thus, it is not necessary to have an operator per vehicle. Currently there are few systems that solve these path planning problems in the literature oriented to agricultural and forestry use, especially dedicated to the optimization of field survey tasks. This sector can be strongly benefited by the group use of aircraft. Therefore, the main field of application of this project is field prospecting.

This objective of this paper is to develop a system for solving the Path Planning problem with 2D grid-based maps adapted to UAVs' sensors with different number of UAVs using Q-Learning techniques.

2. Materials and Methods

This section describes the calculation used for the extraction of the flight maps and the proposed method for the calculation of the flying paths, each described in its corresponding subsection.

2.1. Flight Maps

For the calculation of the flight maps, the cell size is calculated as the projection of the capture area of the sensors on the terrain based on the image size, the flight height and the lens angle of view. In order to better combine the captured data, the smallest area among all UAVs is chosen to take advantage of the overlapping of those with larger capture areas.

No previous information is extracted from the calculated grid-map to direct the calculation of the paths in order to avoid biases. However, by storing also information such as the position of the drones and the cells already visited at each moment, it is possible to provide a great amount of information in real time in order to improve the calculation of the paths.

2.2. Proposed Model

The proposed model for the calculation of the paths is a variation of the Q-Learning algorithm [4]. In this Reinforcement Learning algorithm (RL) [5] the calculation of the q-values is predicted based on an Artificial Neural Network (ANN) [6] with two fully-connected layers with sigmoid activations and the RMSprop optimizer.

To obtain better results in less time, a Hill-Climbing policy [7] is followed to update the rewards received by the UAVs as they move. A training strategy using Memory Replay [8] has also been followed.

Another inherent problem with the proposed models is their configuration with respect to UAVs. There are two possibilities: first, to use a single global ANN for all UAVs; and, second, to use an ANN for each UAV, or local ANN. The first proposal requires less computational resources, but the path calculation for one UAV can be distorted with erroneous information from the paths of the other UAVs. On the other hand, the second approach requires more computational resources, but each ANN is specialized only for each UAV.

3. Results

For the experiments, simulations were carried out in the terrain of the CITIC research center. The metric of interest is the flight time taken to find a solution as it influences the energy consumption of each UAV. Results are listed at Table 1.

Table 1. Table summarizing the best times for each experiment with different numbers of UAVs and ANN configurations.

Number of UAVs	ANN Configuration	
	Global ANN	One ANN per UAV
1 UAV	00:02:19	00:02:19
2 UAVs	00:02:24	00:00:58
3 UAVs	00:01:25	00:01:39
4 UAVs	00:03:00	00:01:13
5 UAVs	00:03:32	00:01:47

4. Conclusions

The calculation of flight path calculation of UAV swarms is approachable by Q-Learning with small full-connected ANNs. This makes the system faster and more efficient than others found in the literature. Thus, facilitating its use by other users. Minimizing the time taken to find each solution is a satisfactory metric that is rarely used by other authors. However, it is one of the most realistic since it is not possible to predict the battery consumption since it depends on other external factors such as the incident wind. One ANN per UAV is usually the best option. As the number of UAVs increases the time taken to find a solution does not grow much more, unlike a global ANN.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: The authors would like to thank the support from NVidia corp., which granted the GPU used in this work. They also acknowledge the support from the CESGA, where many of the preliminary tests were run.

Abbreviations

The following abbreviations are used in this manuscript:

UAV	Unmanned Aerial Vehicle
RL	Reinforcement Learning
ANN	Artificial Neural Network

References

1. Yeaman, M.L.; Yeaman, M. *Virtual Air Power: A Case for Complementing ADF Air Operations with Uninhabited Aerial Vehicles*; Air Power Studies Centre: New Delhi, India, 1998.
2. Zhao, Y.; Zheng, Z.; Liu, Y. Survey on computational-intelligence-based UAV path planning. *Knowl.-Based Syst.* **2018**, *158*, 54–64. [[CrossRef](#)]
3. Campion, M.; Ranganathan, P.; Faruque, S. A review and future directions of UAV swarm communication architectures. In Proceedings of the 2018 IEEE International Conference on Electro/Information Technology (EIT), Rochester, MI, USA, 3–5 May 2018; pp. 0903–0908.
4. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; MIT Press: Cambridge, MA, USA, 2018.
5. Wiering, M.; Van Otterlo, M. Reinforcement learning. *Adapt. Learn. Optim.* **2012**, *12*, 3.
6. Hopfield, J.J. Artificial neural networks. *IEEE Circuits Devices Mag.* **1988**, *4*, 3–10. [[CrossRef](#)]
7. Kimura, H.; Yamamura, M.; Kobayashi, S. Reinforcement learning by stochastic hill climbing on discounted reward. In *Machine Learning Proceedings 1995*; Elsevier: Amsterdam, The Netherlands, 1995; pp. 295–303.
8. Foerster, J.; Nardelli, N.; Farquhar, G.; Afouras, T.; Torr, P.H.; Kohli, P.; Whiteson, S. Stabilising experience replay for deep multi-agent reinforcement learning. International conference on machine learning. *PMLR* **2017**, *70*, 1146–1155.