Reinforcement Q-Learning for Path Planning of Unmanned Aerial Vehicles (UAVs) in Unknown Environments

Article *in* International Review of Automatic Control (IREACO) · November 2023

DOI: 10.15866/ireaco.v16i5.24078

CITATION
1

READS
448

4 authors:

Adam Zourari
Université Hassan 1er
2 PUBLICATIONS **1** CITATION
SEE PROFILE

My abdelkader Youssefi
Université Hassan 1er
31 PUBLICATIONS **80** CITATIONS
SEE PROFILE

Youssef BEN Youssef
Université Hassan 1er
18 PUBLICATIONS **35** CITATIONS
SEE PROFILE

Rachid Dakir
Université Hassan 1er
18 PUBLICATIONS **47** CITATIONS
SEE PROFILE

Praise Worthy Prize

International Review of Automatic Control (I.RE.A.CO.), Vol. 16, N. 5
ISSN 1974-6059 *September 2023*

Reinforcement Q-Learning for Path Planning of Unmanned Aerial Vehicles (UAVs) in Unknown Environments

Adam Zourari[1], My Abdelkader Youssefi[2], Youssef Ben Youssef[3], Rachid Dakir[3]

Abstract – *Path planning for Unmanned Aerial Vehicles in environments with obstacles remains a challenging task. Traditional algorithms, such as A* and Dijkstra, have limitations when dealing with dynamic and changing obstacles, as well as unknown environments. In this paper, a Q-Learning approach for the path planning of UAVs in obstacle-rich and unknown environments is proposed. The impact of alpha, gamma, epsilon, the initial matrix, and reward parameters on the learning process is investigated to achieve safe and cost-effective paths with reduced execution time. The proposed approach is evaluated through simulations by adjusting the alpha, gamma, epsilon, initial matrix, and reward values, and the results demonstrate the effectiveness of the proposed method. The simulation results show that adjusting all the studied parameters can significantly improve the performance of the proposed approach, leading to paths that meet cost and timing objectives while avoiding obstacles.*

Keywords: *Unmanned Aerial Vehicles, Path Planning, Unknown Environments, Q-Learning*

Nomenclature

PID	Proportional Integral Derivative
RADAR	Radio Detection and Ranging
RL	Reinforcement Learning
R	Reward received after taking action
UAVs	Unmanned Aerial Vehicles
d	Distance
α	Learning rate
ε	Exploration rate
γ	Discount factor

I. Introduction

The first use of Unmanned Aerial Vehicles (UAVs) in military operations dates back to the early 20th century, specifically during World War I [1]. The utilization of UAVs during this time has marked a significant advancement in aerial warfare capabilities. The emergence of unmanned aircraft has revolutionized military tactics by providing a means to gather intelligence and conduct surveillance without endangering human lives. These aerial platforms have proved instrumental in obtaining critical information about enemy positions, troop movements, and strategic targets. Additionally, UAVs have enabled scientific exploration and environmental monitoring, allowing researchers to gather valuable data from remote and inaccessible regions [1]. The military applications of UAVs have continued to expand over the years. Modern UAVs are equipped with advanced sensor systems and imaging technologies, enhancing their surveillance capabilities. They have become essential assets in military operations, providing real-time situational awareness, target acquisition, and precision strikes.

Moreover, UAVs have found applications beyond the military realm. They are extensively employed in scientific research for environmental monitoring, mapping, and wildlife conservation efforts including aerial photography, search, exploration [2] and rescue, surveillance, delivery, agriculture [3], and sports [4]. The ability of UAVs to access remote and hazardous areas makes them valuable tools for collecting data and studying various ecosystems. UAVs have numerous benefits, including the ability to operate in complex and hazardous environments, reducing risk, effort, and cost.

However, in order to utilize UAVs effectively, a well-defined and optimized path is essential. The current work focuses on improving the trajectory of UAVs, as well as their navigation and target tracking. [5] The path determines the trajectory that the UAV follows within its operational environment, enabling it to navigate from the starting point to the target destination. A carefully planned and executed path ensures that the UAV can leverage its advantages and achieve its intended objectives. The path may be planned for a single UAV, a swarm of UAVs, or an environment containing other aircraft to avoid collisions with [6]. This path operates in environments that can be known or unknown and may contain obstacles, whether they are mobile or stationary.

Additionally, these environments can be complex and have large dimensions. Therefore, path planning algorithms such as Dijkstra, A*, [1], and D* can work in known and static environments, but they cannot operate in unknown and changing environments. In such environments, [4], where communication networks are

 https://doi.org/10.15866/ireaco.v16i5.24078

unavailable, satellite coverage is absent, or they are subjected to interference [7], remote control is not possible, and transmission of the image is not feasible.

These environments are referred to as unknown environments and are discovered through the use of radar systems, cameras, and sensors [7]. Good path planning deals with the accuracy of sensors, which may have errors [8]. Constructing a conducive environment is an important factor in path planning to provide a collision-free path, as it significantly affects its success and quality [6]. These sensors have a margin of error that should be taken into consideration [8]. This becomes particularly important in unknown environments that are exposed to high wind speeds [9] or involve tasks related to providing communication services in certain areas [10] or with dynamic obstacles. [11] Moreover, not only can the obstacles themselves be dynamic, but the objective can also be dynamic [12]. This necessitates retraining and path planning from the beginning in conventional algorithms. Additionally, good path planning is characterized by minimizing the number of turns, which consume higher energy [12]. Furthermore, aircraft control and navigation may involve inaccuracies and instability, even if aided by PID [13], [14]. These factors should be taken into consideration in path planning. For example, the trajectory may change due to wind variations or alterations in the aircraft's payload.

Therefore, it is imperative to rectify the trajectory, as demonstrated in a study that has employed artificial intelligence to track the trajectory and maintain stability through the utilization of Artificial Neural Networks (ANN) [15]. The changes in wind conditions, along with the instability of the sea surface on a non-aircraft carrier vessel, can result in a shift in the landing location.

Therefore, landing a UAV on such ships requires following the descent axis. This involves developing precise mechanical equations, modeling the appropriate differential equations, and working on achieving stability and consistency in the face of these continuous changes [16]. In one of the studies, automatic calibration for the landing phase at airports has been developed. This phase is considered the most challenging in an aircraft's journey, as it involves collecting data from sensors and cameras to determine altitude and angles for landing procedures and make necessary corrections [17]. The larger the dimensions are, the more time it takes to execute a path change, by taking into consideration that drone microprocessors have limited capabilities unlike large-scale computers [9]. Furthermore, energy consumption in processors and engines is actively worked on to reduce it [18]. In a context characterized by substantial dimensions, path planning demands have escalated computational effort and extended timeframes.

Consequently, the integration of deep learning to predict values that facilitate path planning through the utilization of path planning values via Q-learning is a viable approach [23]. With the advancements in computing and artificial intelligence, it is now possible to work in unknown and variable environments. This is the subject of this work using artificial intelligence, specifically reinforcement learning, and particularly deep reinforcement learning as a preliminary step towards deep learning-based path planning. In the study, the focus has been on identifying the factors of alpha, gamma, epsilon, and reward, along with the addition of an initial matrix as a suitable starting point to achieve an optimal path that minimizes costs by reducing time and iterations, while also avoiding collisions with obstacles.

These obstacles may include objects that should not be collided with or hazardous areas like anti-aircraft radars.

This approach aims to harness the advantages of unmanned vehicles and apply them in unknown environments that entail risks [4]. Path planning [24] is not limited to aircraft but it can also be used in other domains such as underwater submarines [19] or vehicles in exploration areas [20].

This paper has been divided into five sections. The second section explains the application of the q-learning principle and the algorithm employed. In the third section, the results obtained are presented, followed by a detailed discussion in the subsequent section. The fifth section reviews the conclusions and summarizes the research findings and the proposed recommendations

II. Methods

II.1. Introduction to Q-Learning

Q-learning is a subfield of artificial intelligence, specifically within the realm of reinforcement learning. It is a self-learning algorithm used for decision-making based on rewards. The algorithm utilizes a Q-table to evaluate and update values for each state-action pair in a given environment. The objective is to achieve the highest Q-value possible to make optimal decisions [20].

The agent operates in an environment where it takes actions determined for it, and based on those actions, rewards are received and the state changes accordingly.

The quality matrix, which contains actions in its columns and states in its rows, is updated and modified accordingly.

II.2. The Bellman Equations

The algorithm relies on the Bellman equations for updating the Q-table in Q-Learning. These equations represent a balance between the current experience and future reward expectations. The general form of the Bellman equation in Q-Learning [12] is as follows:

$$Q(s,a) = (1-\alpha)Q(s,a) + \\ +\alpha\left(R + \gamma\max\left(Q(s',a')\right)\right) \quad (1)$$

where $Q(s, a)$ represents the Q-value for a particular state (s) and action (a), α is the learning rate, determining the impact of the new update on the current Q-value, R is the reward received after taking action (a) in state (s), γ is

the discount factor, determining the agent's consideration of future rewards, max($Q(s', a')$) represents the highest Q-value for the future state (s') across all possible actions (a'). Additionally, the Q-Learning algorithm involves other parameters: ε is the exploration rate, determining the probability of taking a random action instead of the best-known action from the Q-table. The initial Q-table is a matrix filled with default values before the training process begins

II.3. Implementation Steps

In order to implement Q-Learning, these steps can be followed:

1. Initialize the Q-table with default values;
2. Set values for the parameters (α, γ, ϵ);
3. Determine the number of training episodes (the number of times the environment will be run);
4. Within each training episode:
   - Start from a specific state;
   - Within each step of the episode:
     - Use an action selection policy based on Q-values (exploration or exploitation);
     - Execute the action, calculate the reward, and observe the next state;
     - Update the Q-value using the Bellman equation;
     - Transition to the next state and continue the episode until reaching the goal state.
5. After completing the training process, use the updated Q-table for decision-making based on the current state.

Q-learning has been employed to determine the optimal path in a two-dimensional context, despite drones typically operating in three-dimensional environments. Adapting the algorithm for a three-dimensional setting is feasible. The environment has been represented by using a matrix, where each cell has corresponded to a 10-centimeter square, with specific dimensions assigned to each point based on the matrix's rows and columns. Obstacles have been represented by cells with negative values, while the target cell or goal point had the highest value. Movement between cells has been possible in four directions: up, down, right, and left.

Reaching the goal has resulted in a reward of 20, whereas encountering an obstacle incurred a penalty of -20. Hazardous areas, like the presence of anti-aircraft systems, have been also represented as cells with negative values. In other cases, the reward could vary based on the Euclidean distance. The initial matrix has been populated with values calculated from the Euclidean distance between the current drone location and the target point:

$$d = \sqrt{\left(x_{ij} - x_g\right)^2 + \left(y_{ij} - y_g\right)^2} \tag{2}$$

A lambda value of 100 has been utilized:

$$d_\lambda = e^{\lambda . d} \tag{3}$$

The reward variable has been made dynamic, similar to the table [16]. When the final matrix has been obtained, a comparison between the starting cell and its adjacent cells, including those to the right, left, above, and below, has been performed. The highest value among these cells has determined the next move without revisiting the same point. This process has been repeated until the cell with the highest value, representing the optimal path, has been reached. Rewards have been plotted at each iteration, enabling the observation of the effectiveness of the modifications. The method has been tested in a two-dimensional environment consisting of 10 columns and rows, totalling 100 cells. The start and end points have been defined and kept separate from each other. It has been observed that achieving a good approximation between the start and end points have not required large alpha values. In this particular scenario, the coordinates (0,0) have been set as the starting point, and the endpoint as the goal (7.9). The identified obstacles have been represented as squares and rectangles marked with black symbols, and the path has been indicated by using circles.

III. Result

Case 1: variable rewards have not been used, and the initial matrix values have been set to zero. The values of epsilon and other parameters have been configured as presented in Table II, resulting in the outcome shown in Figure 2. It is noticeable that the system stabilizes after approximately 200 iterations. Case 2: the approach has been similar to the first case regarding rewards and the initial matrix. However, adjustments have been made to the values of alpha and gamma as outlined in Table III.

The resulting outcomes are illustrated in Fig. 3.

TABLE I
REWARD

Setting	Values
Goal	20
Obstacle	-20
Other	1/d

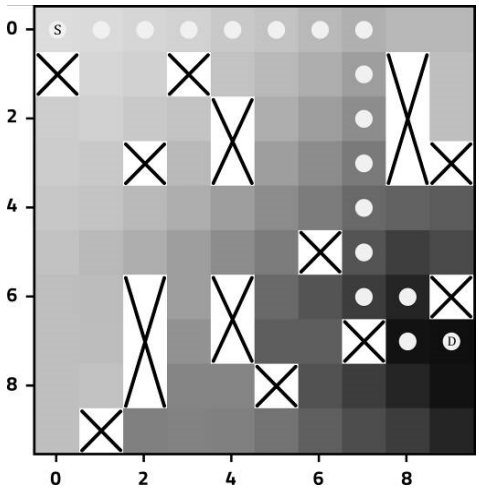

Fig. 1. Path in a binary environment with obstacles

TABLE II
VARIABLE VALUES 1

Setting	Values
Alpha	1
Gamma	0.8
Epsilon	0.01
Iterations	10000

TABLE IIII
VARIABLE VALUES 2

Setting	Values
Alpha	0.2
Gamma	0.5
Epsilon	0.01
Iterations	10000

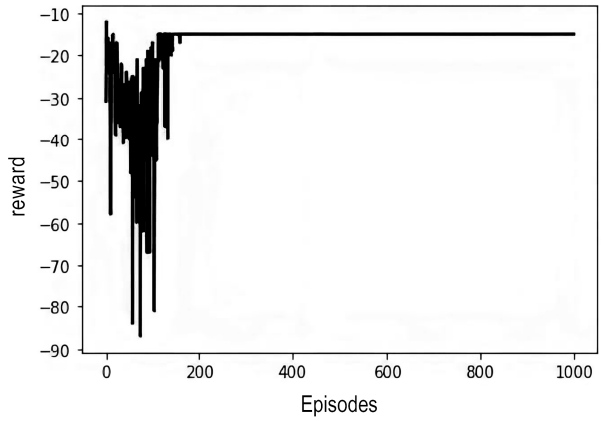

Fig. 2. Reward vs. Episodes Plot 1

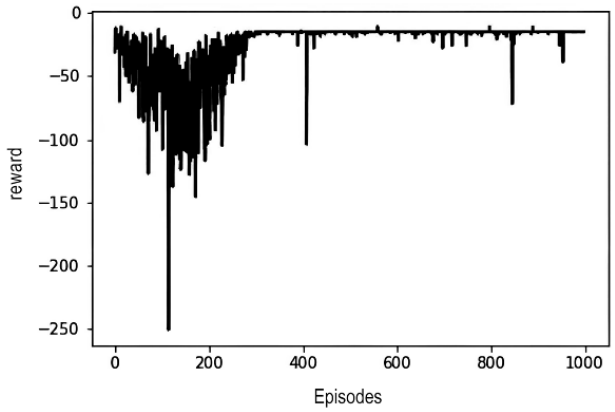

Fig. 3. Reward vs. Episodes Plot 2

Case 3: For this scenario, values of alpha and gamma have been selected based on their best performance in the first case.

Reward values have been adjusted to a constant +1, and the initial matrix has been populated with values corresponding to the Euclidean distance. The results are displayed in Fig. 4. Case 4: For this case, values of alpha and gamma have been selected to match those of the first case. The initial matrix has retained its values associated with the Euclidean distance. However, the reward variable has been adjusted to be inversely proportional to the Euclidean distance. The outcomes are depicted in Fig. 5.

TABLE IV
REWARD 1

Setting	Values
Goal	20
Obstacle	-20
OTHER	1

TABLE V
REWARD 2

Setting	Values
Goal	20
Obstacle	-20
OTHER	1/d

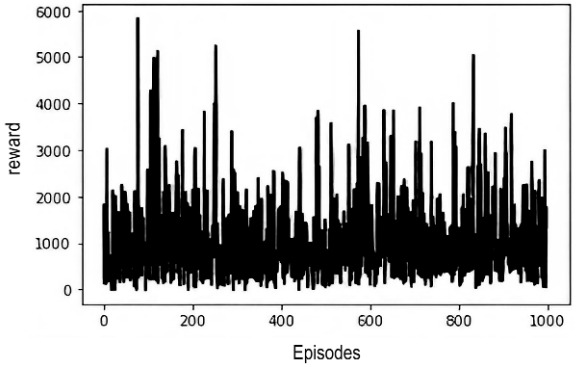

Fig. 4. Reward vs. Episodes Plot 3

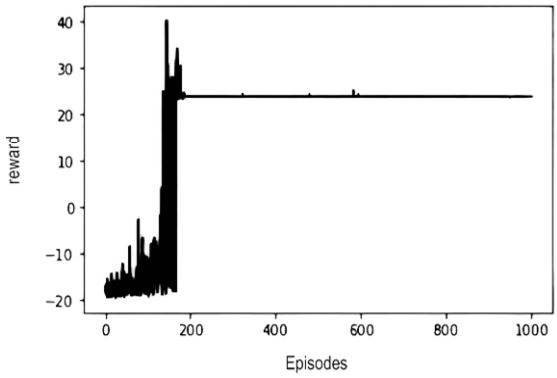

Fig. 5. Reward vs. Episodes Plot 4

IV. Discussion

The results of this study, as revealed through rigorous simulation and program application, has shed valuable insights into the intricate relationship between the learning rate and the path construction for Unmanned Aerial Vehicles (UAVs). It has become unmistakably evident that as the learning rate increases, the efficiency of path construction is significantly improved. This phenomenon can be attributed to the learning rate's profound influence on the weighting of current information in comparison to past information. A higher learning rate signifies an increased emphasis on recent data, which proves exceptionally advantageous. The significance of this finding is particularly pronounced post-exploration, where information becomes increasingly critical compared to its relevance during the initial phases of agent learning. This leads to the

hypothesis that by reducing the epsilon value, the agent to expedite the transition from exploration to exploitation is enabled. This assumption gains further support from the decision to initialize the first quality matrix with values directly related to Euclidean distance. This adjustment effectively compels the agent to exploit these values rather than indulging in exploration. As a result, the reduction of epsilon yields a system that not only maintains stability but also does so without compromising the speed of exploration. It is important to note that this study takes place in a static environment, rendering an extensive exploration rate unnecessary.

These findings have been empirically validated through simulations, especially in scenarios involving variable rewards. In these situations, proximity to the target point translates to larger rewards, further underscoring the efficacy of our approach. Moreover, this study stands out by combining the use of the initial matrix, variable rewards, and dynamic information.

While prior research has often examined variable rewards and individual factors in isolation, this approach is a comprehensive one that integrates these variables, leading to superior results. In this paper, four distinct scenarios have been devised aimed at elucidating the values of alpha, gamma, and epsilon and their impact on the ability of rewards to enhance the path planning process, as quantified by the required iterations for search stability, expressed in terms of reward stability.

The findings are most illuminating in the first and second scenarios, as indicated by the results presented in Figures 2 and 3. In the first scenario, stability has been achieved within a mere 200 iterations, highlighting the success of the approach. However, the third and fourth scenarios, as depicted in Figures 4 and 5, have revealed the failure to attain stability. This underscores the paramount importance of flexible and adaptive reward structures and further emphasizes that maintaining constant rewards does not align with the primary goal of determining the optimal values for alpha, gamma, and epsilon. In this study, a systematic approach has been employed in order to enhance the understanding of the intricate dynamics involved in path planning for Unmanned Aerial Vehicles (UAVs). The presented methodology has involved iteratively adjusting individual parameters while keeping the environmental conditions constant, allowing examining meticulously the influence of each factor. The results not only have highlighted the importance of fine-tuning the learning rate, alpha, gamma, and epsilon but have also emphasized the need to consider their interplay.

Moreover, the observations have indicated that, when working with larger-scale environments, energy consumption and time required for planning significantly have increased. This finding has important implications for real-world applications of UAVs, especially in scenarios where adaptability to changing environments is critical. The presented research lays the foundation for future investigations, which should address the adaptation of models and parameters to dynamic real-world conditions to optimize UAV performance.

V. Conclusion

In this study, an investigation has been conducted to examine the effects of altering the alpha (α), epsilon (ε), and gamma (γ) parameters, as well as the influence of reward and the initial matrix, on the enhancement of pathfinding for Unmanned Aerial Vehicles (UAVs). The findings have demonstrated that modifications to these variables have yielded improved results, highlighting the interconnectedness of these changes for achieving a coherent solution. Furthermore, it has been established that adaptability is crucial when considering different environmental factors and the nature of obstacles, whether they are dynamic or static. In environments characterized by substantial dimensions, the Q-learning algorithm necessitates increased computational effort and time, prompting a potential shift toward the exploration of Deep Q-learning in forthcoming research to enhance further the process. Nevertheless, it is essential to acknowledge that Reinforcement Learning (RL) presents certain limitations when applied to UAVs. These constraints encompass the demand for significant volumes of training data, which can be a resource-intensive endeavor, especially for UAVs with constraints in terms of battery life and flight duration. Safety concerns are also paramount since RL may not guarantee consistent safe operation within complex and unpredictable environments. Moreover, RL agents designed for specific tasks may struggle with generalizing their learned knowledge to novel contexts or scenarios. Striking an equilibrium between exploration and exploitation remains a challenging aspect for RL agents, particularly in UAV applications where exploration can entail certain risks. Additionally, the decision-making process of RL agents may lack transparency, introducing complexities for human operators who require a clear understanding of the agent's actions. Consequently, it is imperative to consider these limitations and consider the integration of a diversified set of techniques to ensure the safe and effective operation of UAVs.

References

[1] S. Aggarwal and N. Kumar, Path planning techniques for unmanned aerial vehicles: A review, solutions, and challenges, *Comput. Commun.* 149, 270–299 (2020). doi: 10.1016/j.comcom.2019.10.014

[2] Szabo, S., Železník, V., Mako, S., Rabatin, R., Kinematics of Exploration Using Unmanned Aerial Vehicles, (2022) *International Review of Aerospace Engineering (IREASE)*, 15 (5), pp. 244-253. doi: https://doi.org/10.15866/irease.v15i5.22361

[3] G. G. d. Castro, G. S. Berger, A. Cantieri, M. Teixeira, J. Lima, A. I. Pereira, and M. F. Pinto, Adaptive path planning for fusing rapidly exploring random trees and deep reinforcement learning in an agriculture dynamic environment uavs, *Agriculture* 13, 354 (2023). doi: https://doi.org/10.3390/agriculture13020354

[4] C. Yan and X. Xiang, A path planning algorithm for uav based on

improved q-learning, in 2018 *2nd international conference on robotics and automation sciences (ICRAS), (IEEE, 2018)*, pp. 1–5.
doi: http://dx.doi.org/10.1109/ICRAS.2018.8443226

[5] M. Shurrab, R. Mizouni, S. Singh, and H. Otrok, Reinforcement learning framework for uav-based target localization applications, *Internet Things* p. 100867 (2023).
doi: http://dx.doi.org/10.1016/j.iot.2023.100867

[6] Y. Cao and X. Fang, Optimized-weighted-speedy q-learning algorithm for multi-ugv in static environment path planning under anti-collision cooperation mechanism, *Mathematics* 11, 2476 (2023).
doi: http://dx.doi.org/10.3390/math11112476

[7] A. El Farnane, M. a. Youssefi, A. Mouhsen, and a. ihyaoui, Visual and lidar-based simultaneous localization and mapping for self-driving cars, *Int. J. Electr. Comput. Eng.* 12, 6284–6292 (2022).
doi: http://dx.doi.org/10.11591/ijece.v12i6.pp6284-6292

[8] R. Alami, H. Hacid, L. Bellone, M. Barcis, and E. Natalizio, Soreo: A system for safe and autonomous drones fleet navigation with reinforcement learning, in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37 (2023), pp. 16398–16400.
doi: https://doi.org/10.1609/aaai.v37i13.27058

[9] J. Yang, S. Lu, M. Han, Y. Li, Y. Ma, Z. Lin, and H. Li, Mapless navigation for uavs via reinforcement learning from demonstrations, *Sci. China Technol. Sci.* pp. 1–8 (2023).
doi: https://doi.org/10.1007/s11431-022-2292-3

[10] A. Souto, R. Alfaia, E. Cardoso, J. Araújo, and C. Francês, UAV path planning optimization strategy: Considerations of urban morphology, microclimate, and energy efficiency using q-learning algorithm, *Drones* 7, 123 (2023).
doi: https://doi.org/10.3390/drones7020123

[11] D. Zhang, Z. Xuan, Y. Zhang, J. Yao, X. Li, and X. Li, Path planning of unmanned aerial vehicle in complex environments based on state-detection twin delayed deep deterministic policy gradient, *Machines* 11, 108 (2023).
doi: http://dx.doi.org/10.3390/machines11010108

[12] N. Ali, K. Kamarudin, M. A. A. Bakar, M. H. F. Rahiman, A. Zakaria, S. M. Mamduh, and L. M. Kamarudin, 2D lidar based reinforcement learning for multi-target path planning in unknown environment, *IEEE Access* (2023).
doi: http://dx.doi.org/10.1109/ACCESS.2023.3265207

[13] M. H. A. Bakar, A. U. bin Shamsudin, R. A. Rahim, Z. A. Soomro, and A. Adrianshah, Comparison method q-learning and sarsa for simulation of drone controller using reinforcement learning, *J. Adv. Res. Appl. Sci. Eng. Technol.* 30, 69–78 (2023).
doi: https://doi.org/10.37934/araset.30.3.6978

[14] Mlayeh, H., Ghachem, S., Nasri, O., Ben Othman, K., Stabilization of a Quadrotor Vehicle Using PD and Recursive Nonlinear Control Techniques, (2021) *International Review of Aerospace Engineering (IREASE)*, 14 (4), pp. 211-219.
doi: https://doi.org/10.15866/irease.v14i4.19739

[15] Housny, H., Chater, E., El Fadil, H., Feedforward Neural Network Controller for Quadrotor in the Presence of Payload and Wind Disturbances, (2021) *International Review of Automatic Control (IREACO)*, 14 (5), pp. 287-299.
doi: https://doi.org/10.15866/ireaco.v14i5.20480

[16] Kramar, V., Alchakov, V., Kabanov, A., Dudnikov, S., Dmitriev, A., The Design of Optimal Lateral Motion Control of an UAV Using the Linear-Quadratic Optimization Method in the Complex Domain, (2020) *International Review of Aerospace Engineering (IREASE)*, 13 (6), pp. 217-227.
doi: https://doi.org/10.15866/irease.v13i6.19130

[17] Uc Ríos, C., Teruel, P., Use of Unmanned Aerial Vehicles for Calibration of the Precision Approach Path Indicator System, (2021) *International Review of Aerospace Engineering (IREASE)*, 14 (4), pp. 192-200.
doi: https://doi.org/10.15866/irease.v14i4.20709

[18] Keserwani, Z., Saied, M., Francis, C., Medium and Low-Level Energy Saving Control Strategies for Electric-Powered UAVs, (2023) *International Review of Automatic Control (IREACO)*, 16 (2), pp. 54-65.
doi: https://doi.org/10.15866/ireaco.v16i2.23120

[19] Z. Wang, H. Lu, H. Qin, and Y. Sui, Autonomous underwater vehicle path planning method of soft actor–critic based on game training, *J. Mar. Sci. Eng.* 10, 2018 (2022).
doi: https://doi.org/10.3390/jmse10122018

[20] J. Jiang, X. Zeng, D. Guzzetti, and Y. You, Path planning for asteroid hopping rovers with pre-trained deep reinforcement learning architectures, *Acta Astronaut*. 171, 265–279 (2020).
doi: http://dx.doi.org/10.1016/j.actaastro.2020.03.007

[21] C. J. Watkins and P. Dayan, Q-learning, *Mach. learning* 8, 279–292 (1992).
doi: http://dx.doi.org/10.1007/BF00992698

[22] Y. Xu, Y. Wei, K. Jiang, D. Wang, and H. Deng, Multiple uavs path planning based on deep reinforcement learning in communication denial environment, *Mathematics* 11, 405 (2023).
doi: https://doi.org/10.3390/math11020405

[23] U. Habiba and R. Jahan, Unmanned aerial vehicle (uav)'drones' using machine learning, Available at *SSRN 4430575* (2023).
doi: https://dx.doi.org/10.2139/ssrn.4430575

[24] Lamini, C., Benhlima, S., Bekri, M., Q-Free Walk Ant Hybrid Architecture for Mobile Robot Path Planning in Dynamic Environment, (2022) *International Journal on Engineering Applications (IREA)*, 10 (2), pp. 105-115.
doi:https://doi.org/10.15866/irea.v10i2.20443

[25] Aguirre, D., Barón Velandia, J., Salcedo Parra, O., Routing in Elastic Optical Networks Based on Deep Reinforcement Learning for Multi-Agent Systems, (2022) *International Review on Modelling and Simulations (IREMOS)*, 15 (5), pp. 332-339.
doi:https://doi.org/10.15866/iremos.v15i5.22768

Authors' information

[1]Laboratory of Engineering, Industrial Management, and Innovation (LIMII) Faculty of Science and Technology,
Hassan First University of Settat,
Morocco.

[2]Interdisciplinary Laboratory of Applied Sciences (LISA) National School of Applied Sciences,
Hassan First University,
Morocco.

[3]Laboratory of Computer Systems and Vision (LabSIV) Polydisciplinary Faculty,
Ibn Zohr University, Ouarzazate,
Morocco.

Adam Zourari, an engineer in Automation, was born in Morocco in 1993. He holds a Bachelor's degree in Mechatronics Engineering. Currently, he is pursuing a doctoral program at Hassan I University in Settat, Morocco. His primary research interests encompass various fields, including programming, path planning, automation, space science, and aviation. His commitment to advancing knowledge in these areas is demonstrated through rigorous research and academic exploration.

Abdelkader Youssefi was born in Tinghir, Morocco. He received his engineering degree in telecommunications from the National Institute of Telecommunications (INPT), Rabat, Morocco, in 2003, and his Ph.D. degree from Mohammadia School of Engineers (EMI) at Mohammed V University in 2015. Dr. Youssefi is currently a teacher at Hassan I University in Settat, Morocco. His research interests include Embedded Systems, self-driving, massive MIMO, wireless communications, and the Internet of Things. He has a strong publication record with more than 20 papers in peer-reviewed journals and referred conference proceeding.

Youssef Ben Youssef holds a Bachelor's degree in electronics from Mohammed 1st University Oujda in 1993 and a Master of Science and Technology in electrical engineering from Hassan 1st University in 2011. He is currently pursuing his PhD Degree in System Analysis and Information Technology Laboratory "ATSI" at Hassan 1st University, Settat, Morocco. His research interests span computer vision, image processing, machine learning, and artificial intelligence. In 1995, he graduated from the aggregation in physics from Higher Normal School. Professor Benyoussef currently holds the position of physics professor in preparatory classes for high schools in Settat, Morocco.

Rachid Dakir was born on January 6, 1982. He obtained his Ph.D. in Network and Telecommunication from the FST University Hassan 1st, Morocco. Dr. Dakir currently serves as a Professor at Ibn Zohr University, FP of Ouarzazate, Morocco, where he is involved in the prototype development of active and passive microwave electronic circuits, systems, networks, IoT, and telecommunications. His research focuses on these areas.