J. Korean Soc. Ind. Appl. Math. Vol.25, No.4, 149–161, 2021 https://doi.org/10.12941/jksiam.2021.25.149

Markov Decision Process-based Potential Field Technique for UAV Planning

CHAEHWAN MOON[1], JAEMYUNG AHN [1†]

[1]DEPARTMENT OF AEROSPACE ENGINEERING, KOREA ADVANCED SCIENCE INSTITUTE OF TECHNOLOGY, DAEJEON, 34141, REPUBLIC OF KOREA
E-mail address: [†]jaemyung.ahn@kaist.ac.kr

ABSTRACT. This study proposes a methodology for mission/path planning of an unmanned aerial vehicle (UAV) using an artificial potential field with the Markov Decision Process (MDP). The planning problem is formulated as an MDP. A low-resolution solution of the MDP is obtained and used to define an artificial potential field, which provides a continuous UAV mission plan. A numerical case study is conducted to demonstrate the validity of the proposed technique.

1. INTRODUCTION

A traditional routing problem framework provides an optimal solution without considering the change in the mission environment. Note the traditional vehicle routing problem (VRP) framework is deterministic and it is difficult to handle actual problem environments involving uncertainty. The MDP can be applied to problems involving stochastic factors (e.g., UAV mission under uncertain environment), addressing the gap in the traditional framework. Various past studies formalized uncertain UAV missions using the MDP framework such as search, rescue, and reconnaissance [1-2]. The use of extended MDP frameworks such as partially-observable MDP is actively studied [3]. A reinforcement learning can be applied to the sequential UAV decision-making process using the MDP formulation [4-6]. Some studies extended the MDP based methodology to the planning of multiple UAVs with the centralized/decentralized approaches [7-9]. Some past studies combined the MDP and the potential field techniques and used the value function map generated by MDP for reinforcement learning with various artificial potential fields [10-11]. The focus of these studies were the online planner for collision avoidance rather than a whole mission planning.

This paper proposes a mission/path planning method for an unmanned aerial vehicle (UAV) based on the artificial potential field (APF) and the Markov decision process (MDP). The proposed method constructs an APF using interpolation of the optimal policies and values obtained from the MDP describing the UAV operations. The APF provides a high-resolution solution (route), which is constructed relatively easily by solving a relatively low-resolution MDP. Hence, the combination of the two approaches enables efficient operations of a UAV with light computing resources.

The remaining parts of this paper are organized as follows. Section 2 introduces the theoretical backgrounds of MDP and UAV mission planning. Section 3 presents the

Received December 5 2021; Accepted December 14 2021; Published online December 25 2021.
2000 *Mathematics Subject Classification*. 90C40.
Key words and phrases. Markov decision process(MDP), Sequential decision-making process, Potential field algorithm, Artificial potential field(APF), Mission planning.
† Corresponding author.

mathematical formulation of a UAV planning problem as an MDP. The artificial potential field technique constructed based on the optimal value function obtained from the MDP is discussed in Section 4. Section 5 discusses the mission planning case study using the proposed method. Finally, Section 6 presents the conclusion of this paper and future study subjects.

2. MARKOV DECISION PROCESS

2.1. **Markov decision process**

The Markov Decision Process (MDP) is a mathematical framework used to model discrete-time probability control processes. It adds rewards and decision-making to the Markov process, a discrete-time stochastic process model. The framework is characterized by the *Markov property* in which the conditional probability distribution to arrive at a future state is affected only by the present state (memorylessness).

$$P(S_t = s_t \mid S_{t-1} = s_{t-1}, \ldots, S_0 = s_0) = P(S_t = s_t \mid S_{t-1} = s_{t-1})$$

where s_t represents the state of an agent at time t. Under this property, the state transition probability is determined only by the current state and the action (A_t) taken. We can also define a reward (R_t) obtained by transitioning to a new state. Thus an agent (decision-maker) changes the state through actions and obtains a reward. As such, the MDP models decision-making in an uncertain situation in which the behavioral results are partially random, and the behavior can only be partially controlled. Figure 1 exhibits the progress of the MDP.

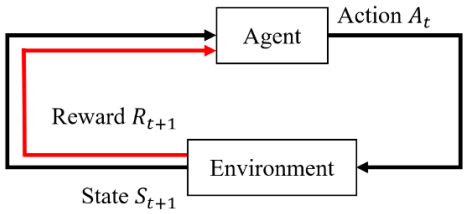

FIGURE 1. The progress of the Markov decision process [17]

The MDP is defined as a 5-tuple (S, A, P, R, γ) whose elements are the state set (S), behavior set (A), state transition probability (P), reward set (R), and depreciation rate (γ). S and A are finite sets of all states/actions. R collects the reward ($R(s, a, s')$) obtainable by selecting action a from current state s to transition to new state s'. Finally, γ converts the value of future rewards into the present value considering uncertainty. The MDP defines a policy function (π) and a value function (V) representing the total reward obtainable by implementing the policy for an optimal decision-making process. π is a probability distribution of action a ($\in A$) for a given state s ($\in S$) representing how a decision-maker behaves defined as follows:

$$\pi(a \mid s) = P[A_t = a \mid S_t = s].$$

The policy can be evaluated using a value function V_π, which is the expected value of the total reward obtained by implementing the policy π. The state-value function $V_\pi(s)$ is the expected value of the total cumulative reward that can be calculated after implementing the policy starting at state s. We define the total cumulative return G_t, which is the sum of rewards that can be obtained after current time *t*, as

$$G_t = R_{t+1} + \gamma R_{t+2} + ... = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}.$$

Then V_π is defined as

$$\begin{aligned} V_\pi &= E_\pi[G_t \mid S_t = s] \\ &= E_\pi\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \mid S_t = s\right]. \\ &= E_\pi\left[R_{t+1} + \gamma G_{t+1} \mid S_t = s\right] \end{aligned}$$

An optimal policy $\pi^*(s)$ is the behavior that enables the transition from the current state (*s*) to the state with the highest value function expressed as

$$\pi^*(s) = \arg\max_a V_\pi(s'),$$
$$V^*(s) = \arg\max_\pi V_\pi(s).$$

2.2. **Bellman Equation**

Bellman equation solves the value function $V_\pi(s)$ using a recursive relationship, which yields the optimal policy $\pi^*(s)$ that maximizes the value function. The Bellman expectation equation is expressed as follows

$$\begin{aligned} V_\pi(s) &= E_\pi\left[\sum_{t=0}^{\infty} \gamma^t R(s_t)\right] \\ &= E_\pi\left[R_{t+1} + \gamma G_{t+1} \mid S_t = s\right] \\ &= E_\pi\left[R_{t+1} + \gamma V_\pi(S_{t+1}) \mid S_t = s\right] \\ &= \sum_{a \in A} \pi(a \mid s)\left(R_s^a + \gamma \sum_{s' \in S} p(s' \mid s, a) V_\pi(s')\right) \end{aligned}$$

In the equations, R_{t+1} is the present immediate reward and $\gamma V_\pi(S_{t+1})$ is the discounted value of the successor state. The optimal state value function for the optimal policy π^* should have the highest value. The value iteration technique determines the optimal policy by using the Bellman Optimality Equation [12] expressed as follows

$$V_*(s) = \max_a r(a,s) + \gamma \sum_{s' \in S} p(s' \mid s,a) V_\pi(s')$$
$$= \max_a \left\{ R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a V_\pi(s') \right\} \quad ,$$

$$V_{k+1}(s) = \max_{a \in A} \left\{ R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a V_\pi(s') \right\} \quad .$$

Note that the MDP is a P-complete problem that satisfies the global optimality condition [13-14]. The computational complexity of an infinite-horizon, discounted-reward MDP used in this study is $O(MN^2)$. Because the worst-case number of iterations is proportional to $-\log(1-\gamma)/(1-\gamma)$, the algorithm converges if γ is not equal to 1 [15].

3. MDP-BASED UAV MISSION PLANNING FORMULATION

3.1. **UAV Mission problem**

3.1.1. **Baseline UAV mission model**

The reference mission discussed in this study is a single UAV mission (visiting multiple target locations and returning to the base) under the presence of threats. The mission may include tasks such as delivery, surveillance, close-air support (CAS), and intelligence, surveillance, and reconnaissance (IS&R). To simplify the problem, we assumed a two-dimensional motion of the UAV at a constant speed without change in the altitude. The UAV aims to visit $N_{mission}$ mission locations and return to one of N_{base} bases. The fuel consumption is proportional to the travel distance. The fuel shortage during the flight leads to the mission failure and refueling at bases is possible. Different weights are assigned to target locations representing their importance. There are N_{threat} hostile threats (radar) against the UAV. The risk level is modeled as a function depending on the distance between the UAV and the threat location [2]. Figure 2 illustrates the UAV mission considered in the study.

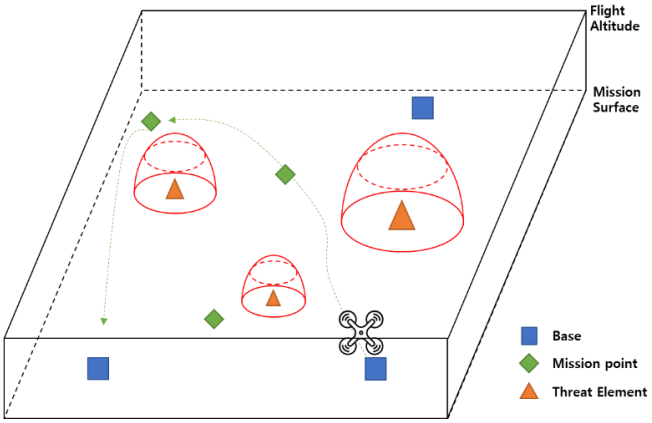

FIGURE 2. UAV mission considered in the study [17]

3.1.2. **Modeling components within the mission environment**

We model the risk associated with the threat as the failure probability (due to the shoot-down) depending on the distance between the UAV and the threat location as follows [16]

$$p_{fail,i} = 1/(1+(r_i/c)^4), \tag{3.1}$$

$$p_{fail} = 1-\prod_{i=1}^{n}(1-p_{fail,i}), \tag{3.2}$$

where r_i is the distance between the UAV and the ith threat, $p_{fail,i}$ is the shoot-down probability due to the i^{th} threat, and p_{fail} is the overall failure probability caused by the threats. Equation (3.1) reflects the fact that the probability of detection increases when the UAV gets closer to the threat (radar), where constant c indicates transmitted power based on two-ray ground-reflection model. Equation (3.2) expresses the overall mission failure probability. The mission success is finally achieved when the UAV reaches the goal location.

3.2. **Problem formulation**

3.2.1. **State set (*S*)**

The state set S is defined as follows

$$S \in \{\mathbf{s} \mid \mathbf{s} = [\mathbf{q}, h, f, \delta_1, ..., \delta_n, \Delta]\}.$$

The components of a state are the location (**q**), health (h), fuel (f), and task completion statuses (δ_i). A location (**q**) has two elements (x and y elements) and is defined on two-dimensional grids. The health (h) represents is a Boolean variable whose value is 1 if the UAV is operational and 0 otherwise (fuel shortage or shoot-down). The fuel (f) represents the amount of fuel remaining during the mission. It can take an integer value ranging between 1

and $f_{\max}$ (100 % fuel). δ_i is a Boolean variable whose value is 1 if task i is completed and 0 otherwise. Finally, Δ is a Boolean variable whose value is 1 if the whole mission is completed and 0 otherwise.

3.2.2. **Action set A**

Set A collects the actions that the UAV can take defined as

$$A = \{a_{move,d}, a_{task,i}, a_{refuel}, a_{ter}\}. \tag{3.3}$$

In Eq. (3.3), $a_{move,d}$ is the movement action capable of moving the UAV in direction d (up, down, left, and right). $a_{task,i}$ completes the i^{th} task. Note that this action is admissible only at the i^{th} task location. a_{refuel} is the refueling action, which is admissible at one of bases. The action fully replenish the fuel ($f = f_{\max}$). Finally, a_{ter} terminates the mission. It is admissible when all the tasks have been completed and the UAV has returned to one of bases.

3.2.3. **State transition probabilistic model P**

The state transition probability ($P(\mathbf{s}'|\mathbf{s},a)$) representing the probability that the current state (**s**) transitions to the next state (**s**') by an action (a) is expressed as follows

$$\begin{aligned} P(\mathbf{s}'|\mathbf{s},a) &= P([\mathbf{q}', h', f', \delta_1', \cdots, \delta_n', \Delta]|\mathbf{s},a) \\ &= P(\mathbf{q}'|\mathbf{s},a)\cdot P(h'|\mathbf{s},a)\cdot P(f'|\mathbf{s},a)\cdot P(\delta_1'|\mathbf{s},a)\cdot\cdots\cdot P(\delta_n'|\mathbf{s},a)\cdot P(\Delta'|\mathbf{s},a) \end{aligned}.$$

The UAV position change occurs deterministically, leading to the following transition probability

$$P(\mathbf{q}'|s,a) = P(\mathbf{q}'|\mathbf{q},a) = \begin{cases} 1 & ((\mathbf{q},a,\mathbf{q}') \text{ is admissible}) \\ 0 & (\text{otherwise}) \end{cases}. \tag{3.4}$$

In Eq. (3.4), the movement is admissible if **q**' is reachable from **q** by action a. We stochastically modeled the transition in the health state h as follows

$$P(h'|\mathbf{s},a) = P(h'|h,\mathbf{q},f,a) = \begin{cases} 1-p_{fail}(\mathbf{q},a) & (h=1, h'=1, f>1) \\ p_{fail}(\mathbf{q},a) & (h=1, h'=0, f>1) \\ 1 & (h=1, h'=0, f=1) \\ 1 & (h=0, h'=0) \\ 0 & (\text{otherwise}) \end{cases}. \tag{3.5}$$

In Eq. (3.5), p_{fail} is defined in Eqs. (3.1)-(3.2). The fuel state transition is deterministically modeled as

$$P(f'|\mathbf{s},a)=P(f'|f,\mathbf{q},a)=\begin{cases}1 & (f'=f-1,\ a=a_{mission}\ \text{or}\ a_{move})\\ 1 & (f'=f_{\max},\ l=l_{base},\ a=a_{refuel})\\ 0 & (\text{otherwise})\end{cases}.$$

A task completion state transition is deterministically modeled as follows:

$$P(\delta_i'|\mathbf{s},a)=P(\delta_i'|\delta_i,\mathbf{q},a)=\begin{cases}1 & (\delta_i'=1,\delta_i=0,\mathbf{q}=\mathbf{q}_{task,i},a=a_{task,i})\\ 1 & (\delta_i'=0,\delta_i=0,\mathbf{q}\neq\mathbf{q}_{task,i}\ or\ a\neq a_{task,i})\\ 0 & (\text{otherwise})\end{cases}.$$

Finally, completion of the whole mission is modelled deterministically as follows:

$$P(\Delta'|\mathbf{s},a)=P(\Delta'|\mathbf{q},\delta_1,\cdots,\delta_n,a)=\begin{cases}1 & (\Delta'=1,\Delta=0,\delta_1\times\cdots\times\delta_n=1,\mathbf{q}=\mathbf{q}_{base},a=a_{ter})\\ 1 & (\Delta'=0,\Delta=0,\delta_1\times\cdots\times\delta_n=0\ or\ \mathbf{q}\neq\mathbf{q}_{base}\ or\ a\neq a_{ter})\\ 0 & (\text{otherwise})\end{cases}.$$

3.2.4. **Reward model R**

The compensation function $R(\mathbf{s},a,\mathbf{s}')$ models the gain/loss caused by state transitions as

$$R(\mathbf{s},a,\mathbf{s}')=R_{mission}(\mathbf{s},a,\mathbf{s}')+R_{failure}(\mathbf{s},a,\mathbf{s}')+R_{fuel}(\mathbf{s},a,\mathbf{s}'),$$

where $R_{mission}$ is a positive reward for mission success, $R_{failure}$ is a negative reward for mission failure caused by shoot-down or fuel exhaustion, and R_{fuel} is a negative reward for fuel consumption. k_{fuel} is a value setting to control of fuel loss aspect, 0.0001 was used. The reward (loss) elements are modeled as

$$R_{mission}=\begin{cases}R_{task,i} & (\delta_i=0,a=a_{task,i},\delta_i'=1)\\ 0 & (\text{otherwise})\end{cases},$$

$$R_{failure}=\begin{cases}-100 & (h=1,h'=0)\\ 0 & (\text{otherwise})\end{cases},$$

$$R_{fuel}=\begin{cases}-k_{fuel} & (a\neq a_{refuel})\\ 0 & (\text{otherwise})\end{cases}.$$

Note that these reward elements are designed so that the UAV performs its mission safely while selecting a route to consume minimum amount of fuel in the mission environment.

4. MDP VALUE BASED ARTIFICIAL POTENTIAL FIELD

4.1. **Potential field path planning method**

The potential field path planning technique designs potential functions associated with the target and the obstacles representing the attracting and repulsive forces, respectively. The potential function considering these forces are presented as

$$U(\mathbf{q}) = \sum_i U_{att,i}(\mathbf{q}) + \sum_j U_{rep,j}(\mathbf{q}) .$$

In the equation, U_{att} and U_{rep} are the potential function components representing the attractive and repulsive forces defined as

$$U_{att,i}(\mathbf{q}) = \frac{1}{2}\xi d^2(\mathbf{q}, \mathbf{q}_{goal,i}) ,$$

$$U_{rep,j}(\mathbf{q}) = \begin{cases} \frac{1}{2}\xi\left(\frac{1}{d(\mathbf{q}, \mathbf{q}_{obs,j}) - Q}\right)^2 & (d(\mathbf{q}, \mathbf{q}_{obs,j}) > Q) \\ 0 & (d(\mathbf{q}, \mathbf{q}_{obs,j}) \le Q) \end{cases} .$$

where $\mathbf{q}$, $\mathbf{q}_{goal,i}$, and $\mathbf{q}_{obs,j}$ are the locations of the agent, the i^{th} goal, and the j^{th} obstacle, respectively. ξ is positive constant scaling factor for making suitable potential field and also Q is positive constant that adjusts effective distance of obstacle's potential field. $d(\mathbf{q}, \mathbf{r})$ is the Euclidean distance between two positions ($\mathbf{q}$ and $\mathbf{r}$). The attractive and repulsive forces are expressed by taking the gradients of the potential functions are follows

$$\begin{aligned} \mathbf{F}_{att}(\mathbf{q}) &= -\nabla U_{att}(\mathbf{q}) = -\nabla \frac{1}{2}\xi d^2(\mathbf{q}, \mathbf{q}_{goal}) \\ &= -\frac{1}{2}\xi(2d(\mathbf{q}, \mathbf{q}_{goal}))\nabla d(\mathbf{q}, \mathbf{q}_{goal}) \\ &= -\xi(\mathbf{q} - \mathbf{q}_{goal}) \end{aligned} ,$$

$$\begin{aligned} \mathbf{F}_{rep}(\mathbf{q}) &= -\nabla U_{rep}(\mathbf{q}) = -\nabla\left(\frac{1}{2}\xi\left(\frac{1}{d(\mathbf{q}, \mathbf{q}_{obs})} - \frac{1}{Q}\right)^2\right) \\ &= \xi\left(\frac{1}{d(\mathbf{q}, \mathbf{q}_{obs})} - \frac{1}{Q}\right)\left(\frac{1}{d^2(\mathbf{q}, \mathbf{q}_{obs})}\right)\frac{\mathbf{q} - \mathbf{q}_{obs}}{\|\mathbf{q} - \mathbf{q}_{obs}\|} \end{aligned} .$$

When the entire potential field is constructed, the potential-guided path planning is performed to move the agent based on the attractive/repulsive forces.

4.2. **MDP based artificial potential field**

A traditional potential field-based path planning approach introduced in the previous subsection moves the agent toward the direction of potential field gradient. However, the traditional method may generate local minima for non-convex obstacle shapes and the quality of the plan for instances with multiple target points are sometimes not very high. This paper addresses this issue by constructing a new artificial potential field based on the values obtained by the (low-resolution) MDP. In MDP, it is the optimal decision to take an action toward a high value, so if this value is reversed and used as a potential field, the gradient at a specific point can be said to be similar to the optimal decision in MDP. As an illustrative

example, Figure 3 presents the spatial distribution of value with a given state (other than location) obtained by the MDP. Figure 4 shows the distribution of potential constructed based on the values set up with the MDP solution (Figure 3). The value obtained from MDP were reversed to point lowest point of new potential field, similar as selecting optimal direction in MDP value distribution. Note that the values are obtained in a relatively coarse grid system, however, the potential generated using the values by interpolation can generate the path in a continuous space.

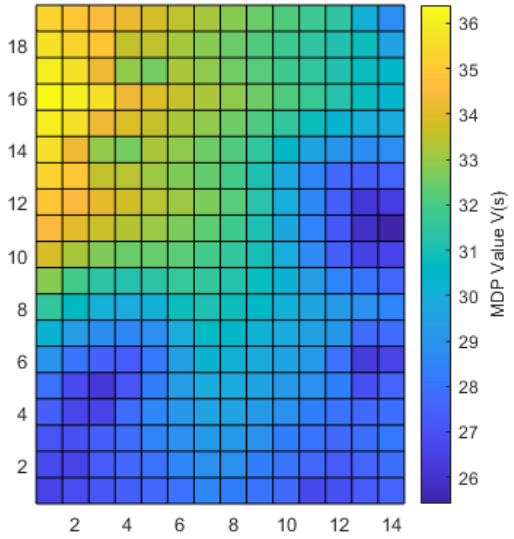

FIGURE 3. Spatial distribution of value obtained by solving the MDP [17]

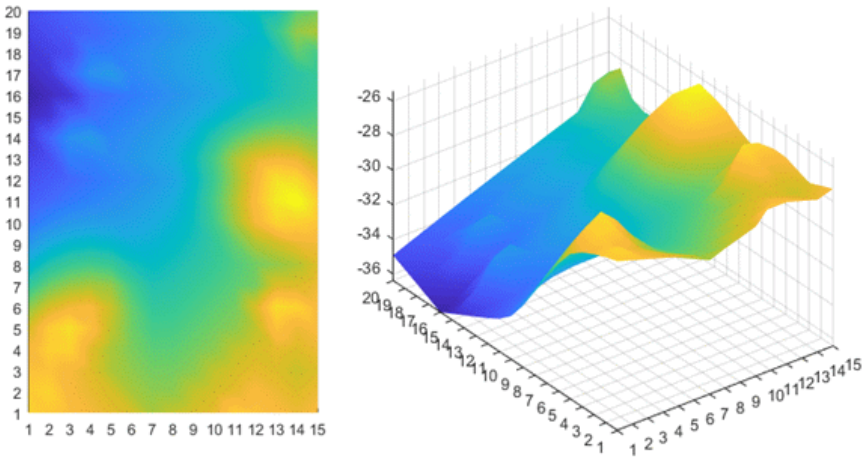

FIGURE 4. Artificial potential field constructed based on MDP value [17]
(interpolated & reversed MDP value to imitate optimal policy)

5. CASE STUDY

This section presents a UAV planning case study using the MDP-based potential field technique introduced in this paper. The reference mission contains 3 targets (tasks), 2 bases, and 6 threats. Figure 5 shows the 2-D map illustrating the mission elements. Tables 2 and 3 summarizes the parameter setting used for the case study.

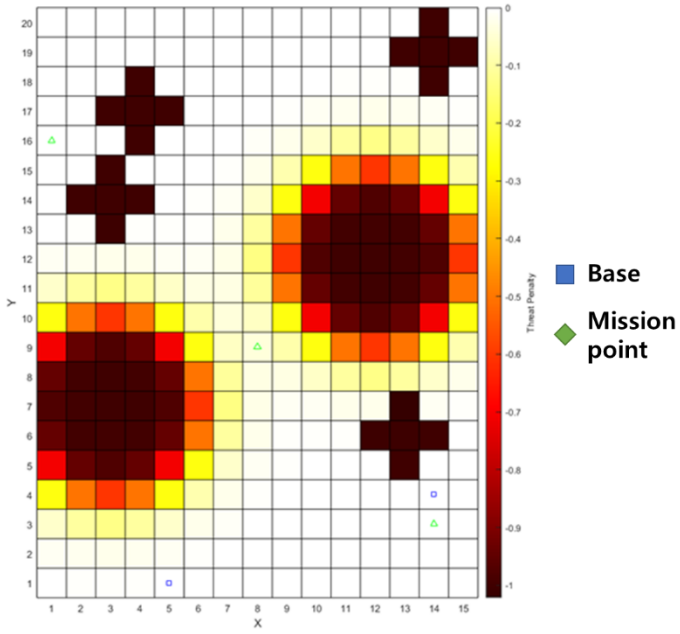

FIGURE 5. UAV mission elements for Case Study [17]

TABLE 1. MDP parameters for Case Study

Parameter	γ (-)	ò	Max. No. of Iteration (-)
Value	0.99	0.001	2000

TABLE 2. Tasks reward and locations for the Case Study

Task	Task Reward ($R_{task,i}$) (-)	Location ($\mathbf{q}_{task,i}$)
Task 1	10	(8, 9)
Task 2	30	(1, 16)
Task 3	10	(14, 3)

TABLE 3. Threat locations for the Case Study

Threat	Threat Location ($\mathbf{q}_{threat,i}$)
Threat 1	(3,7)
Threat 2	(12,12)
Threat 3	(4,17)
Threat 4	(13,6)
Threat 5	(3,14)
Threat 6	(14,19)

5.1 **Case 1: Potential field update upon task-completion**

We conducted a UAV path planning for the Case Study instance using the proposed MDP-based potential field technique. Figure 6 shows the UAV path and associated change in the potential field throughout the mission period. All the three tasks are successfully conducted in the order of tasks 1, 2, and 3. Note that the potential field is updated with the change in the non-spatial state variables, which is shown in the right subfigure of Fig. 6. For example, once Task 1 is completed, the MDP solution lowers the value of moving to the completed task, which is reflected in the updated potential field constructed based on the MDP solution.

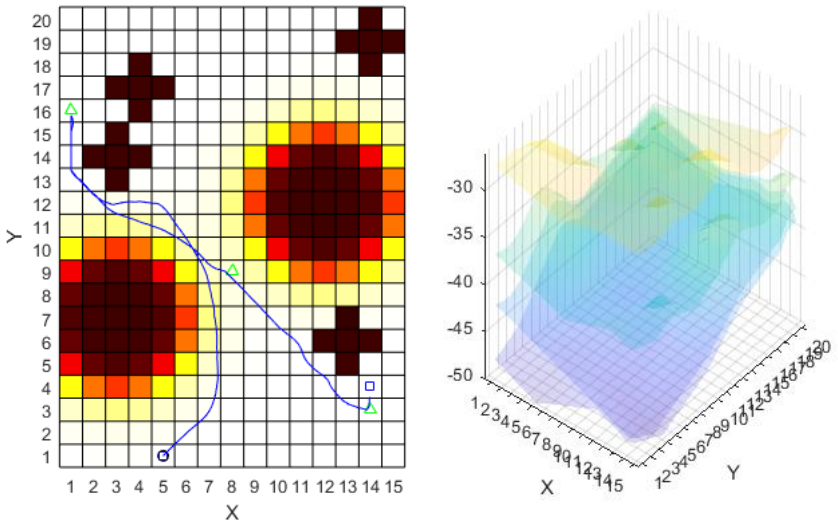

FIGURE 6. Case 1 results: UAV trajectory (Left), Potential field change (Right) [17]

5.2 **Case 2: Potential field update upon remaining fuel change**

In the second case, the potential field is updated upon change in the remaining fuel. Figure 7 shows the UAV path and associated change in the potential field throughout the case 2 mission period. As a result, the potential fields has been updated 48 times (total 49 potential fields used) during the mission. The order of task completion is Task 1 → Task 2 → Task 3

and the UAV trajectory was very similar to the result of Case 1. However, we observed significant change in the potential field which is shown in the right subfigure of Fig 7, when the fuel level is too low to complete the mission without refueling. In this case, the planner commands the UAV visit the vase for refueling instead of conducting additional tasks.

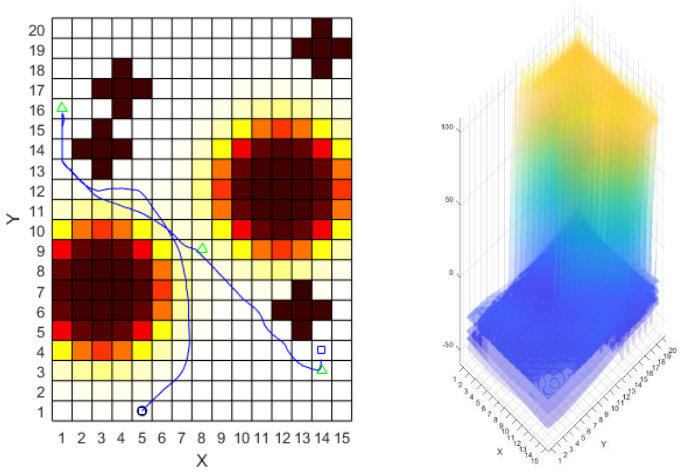

FIGURE 7. Case 2 results: UAV trajectory (Left), Potential field change (Right) [17]

6. CONCLUSION

A UAV mission planning technique using the artificial potential field (APF) constructed based on the (low-resolution) Markov decision process (MDP) is proposed in this paper. In the proposed approach, the path of a UAV is generated based on the attractive/repulsive forces caused by the potential fields associated with the targets/obstacles, which are determined using solution of the MDP. MDP formulation for UAV mission planning composed of multiple tasks, refueling bases, and threats is introduced and the technique to use its solution for potential field construction is presented. A UAV mission planning case study using the proposed technique demonstrates the efficacy of the approach.

Study on the applicability of the proposed approach for various mission types can be one of interesting future research subject. By adding altitude-related states and action, mission environment can be expanded in three dimensions as well. An extension of the proposed approach to multi-agent problems can be also meaningful subject for future study.

ACKNOWLEDGMENTS

This paper is based on the master's thesis of the first author (C. Moon) [17], which was originally written in Korean.

REFERENCES

[1] Waharte, S., & Trigoni, N., Supporting search and rescue operations with UAVs. In 2010 International Conference on Emerging Security Technologies, IEEE, 2010, pp. 142-147.

[2] U. Choi, S. Jeong, J. Ahn, *Autonomous Single UAV Reconnaissance Mission Planning in Multi-Base and Multi-Threat Environment Based on Markov Decision Process*, 2016 KSAS Fall Conference, Jeju, Korea 2016.

[3] S chesvold, D., Tang, J., Ahmed, B. M., Altenburg, K., & Nygard, K. E. *POMDP planning for high level UAV decisions: Search vs. strike*. In In Proceedings of the 16th International Conference on Computer Applications in Industry and Engineering, 2003.

[4] Ure, N. K., Chowdhary, G., Chen, Y. F., How, J. P., & Vian, J. *Distributed learning for planning under uncertainty problems with heterogeneous teams. Journal of Intelligent & Robotic Systems*, 74(1-2) (2014), 529-544.

[5] Lei, G., Dong, M. Z., Xu, T., & Wang, L. *Multi-agent path planning for unmanned aerial vehicle based on threats analysis*. In 2011 3rd International Workshop on Intelligent Systems and Applications, IEEE, 2011, pp. 1-4.

[6] Challita, U., Saad, W., & Bettstetter, C., Deep reinforcement learning for interference-aware path planning of cellular-connected UAVs. In 2018 IEEE International Conference on Communications (ICC) IEEE, 2018, pp. 1-7.

[7] Bethke, B., Redding, J. and How, J. P., *Agent Capability in Persistent Mission Planning using Approximate Dynamic Programming*, 2010 American Control Conference, 2010.

[8] B. Jeong G. Kim ,J. Ha , H. Choi., *MDP based Mission Planning for multi-agent information gathering*, 2013 KSAS Fall Conference, Jeju, Korea 2013.

[9] Jeong, B. M., Ha, J. S., & Choi, H. L. *MDP-based mission planning for multi-UAV persistent surveillance*. In 2014 14th International Conference on Control, Automation and Systems, ICCAS 2014, IEEE, 2014, pp. 831-834.

[10] Bhowal, A. *Potential Field Methods for Safe Reinforcement Learning: Exploring Q-Learning and Potential Fields*. Master's thesis, TU Delft, Delft, Netherlands, 2017.

[11] Zeng, J., Ju, R., Qin, L., Hu, Y., Yin, Q., & Hu, C., *Navigation in Unknown Dynamic Environments Based on Deep Reinforcement Learning*. Sensors, 19(18) (2019), 3837.

[12] Bellman, R., *A Markovian decision process*. Journal of mathematics and mechanics, (1957), 679-684.

[13] Shapley, L. S., *Stochastic games*. Proceedings of the national academy of sciences, 39(10) (1953), 1095-1100.

[14] Papadimitriou, C. H., & Tsitsiklis, J. N., *The complexity of Markov decision processes*. Mathematics of operations research, 12(3) (1987), 441-450.

[15] Littman, M. L., Dean, T. L., & Kaelbling, L. P., *On the complexity of solving Markov decision problems*. In Proceedings of the Eleventh conference on Uncertainty in artificial intelligence, 1995, pp. 394-402.

[16] Jakes, W. C., & Cox, D. C., *Microwave mobile communications*. Wiley-IEEE Press, 1994.

[17] Moon, C., UAV Mission Planning Using MDP-based Artificial Potential Field, Master's Thesis, Korea Advanced Institute of Science and Technology (KAIST), 2021 (written in Korean).