

탑재체를 탑재한 조종 낙하산의 심층 강화학습을 통한 자동 착륙

박준우*, 방효충

한국과학기술원

Automatic Landing of Loaded Parafoil using Deep Reinforcement Learning

Junwoo Park*, Hyochoong Bang

Key Words : Parafoil(조종 낙하산), Automatic Landing(자동 착륙), Deep Reinforcement Learning(심층 강화 학습), Markov Decision Process Design(마르코프 결정 과정 설계)

서 론

낙하산(Parachute)은 추가적인 동력원이 없는 상황에서 적은 추가비용 및 구조물만으로 동체에 설치할 수 있고, 캐노피의 높은 항력을 이용해 탑재체의 낙하 및 착륙 속도를 효과적으로 줄일 수 있는 특징으로 인해 화성 탐사 로버 Perseverance 호의 진입 및 하강 속력의 감소, 우주 임무로부터 지구로 복귀하는 유인 캡슐의 연착륙 및 회수에 활용⁽¹⁾하는 등 큰 규모의 항공우주 시스템에 적용된 사례가 많다. 낙하산의 사용은 공력이 지배적으로 작용할 수 있는 근지구 환경에서 특히 효과적이지만 동시에 바람과 같은 외란의 영향을 많이 받기 때문에 정밀 착륙이 요구되는 재사용 발사체의 경우에는 동력 하강의 형태로⁽²⁾ 회수하기도 한다. 하지만 이는 개발 및 운용 비용이 막대하다.

조종 낙하산(Parafoil)은 캐노피 부분이 여러 익형을 이어붙인 형태로 좌우 양 끝에 조종 가능한 타면이 존재해서 나선형의 궤적을 그리면서 긴 체공 시간 동안 느린 속도로 선회 하강하는 것이 가능하다. 조종 명령을 통해 특정 궤적을 추종하거나 변화하는 주변 상황에 대응할 수 있으므로 일반 낙하산을 통한 착륙보다 활용도가 높다고 할 수 있다. 하지만 충실도(Fidelity) 높은 동역학 모델을 확보하기가 어렵고, 여러 부정확한 공력계수들에 의존하기 때문에 이를 이용한 정밀 착륙은 매우 도전적인 문제이며 바다, 사막 혹은 숲과 같이 회수가 어려워 착륙을 지양해야 하는 곳으로 유도되기 쉽다.

본 연구는 상기와 같이 큰 항공우주 시스템의 효과적인 회수를 위해 조종 낙하산을 활용한 연착륙을 상정한다. 특히, 동력 하강 방식의 정밀 착륙에 준하는 성능을 확보하기 위한 기초 연구로서 특정 형태의 지상 장애물을 회피하고 연착륙하는 과정을 조종 낙하산의 동역학 관계 및 주변 환경의 천이 모델이 전무한 상황에서 강화학습을 이용해 도출하고자 한다.

본 론

조종 낙하산의 캐노피(p)와 탑재체(b)가 질량이 없는 중간 연결 부위(c)를 기준으로 특정 거리만큼 떨어진 채 강체 연결되어 있다는 가정으로 각각의 병진 운동과 회전 운동을 기술하는 뉴턴-오일러 방정식을 이

용하면⁽³⁾ 식 (1)과 같은 9자유도 운동 방정식을 유도할 수 있다.

$$\begin{bmatrix} -M_b R_b & 0 & M_b C_b^m & C_b^m \\ 0 & -M_p R_{cp} & M_p C_p^m & -C_p^m \\ I_b & 0 & 0 & -R_{cb} C_b^m \\ 0 & I_p + I_F & 0 & R_{cp} C_p^m \end{bmatrix} \begin{bmatrix} \dot{\Omega}_b \\ \dot{\Omega}_p \\ v_c \\ F_c \end{bmatrix} = \begin{bmatrix} B_1 \\ B_2 \\ B_3 \\ B_4 \end{bmatrix} \quad (1)$$

여기서 $B_i (i=1\sim 4)$ 는 각각 병진 운동(p, b 순), 회전 운동(p, b 순)과 관련된 항으로 식 (2)를 만족한다.

$$\begin{aligned} B_1 &= F_b^A + W_b - \Omega_b \times M_b \Omega_b \times R_{cb} \\ B_2 &= F_p^A + W_p - \Omega_p \times (M_p + M_F) \Omega_p \times R_{cp} \\ &\quad + M_F \Omega_p \times C_p^m \\ B_3 &= -\Omega_b \times I_b \Omega_b \\ B_4 &= M_p^A - \Omega_p \times (I_p + I_F) \Omega_p \end{aligned} \quad (2)$$

I , M 은 각각 관성 능률, 대각 행렬로 변환한 질량이고 아래 첨자 F 는 겉보기(Apparent) 물성치에 해당하는 것으로 M_F , I_F 는 각각 조종 낙하산 캐노피의 겉보기 질량, 겉보기 관성 능률을 의미한다. 유체 내에서 가속하는 강체의 겉보기 질량 혹은 동적 질량에 대한 상세 내용은 Lissaman 등⁽⁴⁾의 연구결과를 참고할 수 있다. F 와 M 은 각각 힘과 모멘트 벡터, W 는 무게, 상첨자 A 는 공력 항을 나타낸다. M 과 M 이 다른 것에 유의하자. R 과 Ω 는 첨자에 맞는 모멘트 암과 각속도 벡터의 교대대칭(Skew symmetric) 행렬이다. 아울러, 상첨자 A 로 표기된 공력 항에는 식 (3)으로 모델링할 수 있는 조종 낙하산 좌우의 조종면 타각 δ 이 포함되어 있다. ($0 \leq \delta_i \leq 1, \delta_i \in R, i = L, R$) 공력계수와 그 관계식, 식 (1), 그리고 기타 운동학적 관계의 상세 내용은 Prakash 등⁽³⁾의 연구내용을 참고한다.

$$\delta = \begin{bmatrix} \delta_L \\ \delta_R \end{bmatrix} \equiv \begin{bmatrix} \delta_s \\ \delta_a \end{bmatrix} = \begin{bmatrix} \min(\delta_L, \delta_R) \\ \delta_R - \delta_L \end{bmatrix} \quad (3)$$

본 연구에서는 강화학습 환경에 식 (1),(2)의 동역학을 가진 조종 낙하산을 포함해 식 (3)의 조종 명령을 Action으로 설계한다. 또한, 강화학습 Agent는 해당 조종 명령을 사용해 지상에 있는 모델링된 장애물을 회피하여 연착륙하는 정책을 환경과의 상호작용을 통

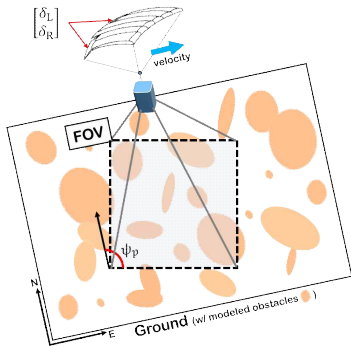


Fig. 1. Depiction of parafoil landing problem onto ground obstacles

해 학습한다. Fig. 1에는 전체적인 문제 상황을 도시하였다.

직접적인 동역학 관계를 모르는 Agent가 조종 낙하산과 지상 장애물 상태를 포함한 문제 상황을 잘 추론하기 위해서는 적합한 관측값 설계가 필수적이다. 시야 안에 포함된 장애물의 개수는 시간에 따라 변할 수 있으므로 이들의 모델링 파라미터들을 단순 나열한 특성 벡터는 차원이 시변한다. 하지만 Actor로 활용되는 심층 신경망은 일반적으로 MLP의 형태로 고정된 크기의 입력만을 수용할 수 있어서 고정된 크기의 관측값을 가진 환경을 설계하는 것이 일반적이다. CNN은 가변적인 크기의 이미지를 입력으로 받을 수 있지만 출력단인 Action의 차원이 정해져 있는 본 문제에서는 마찬가지로 고정된 크기의 입력을 요구한다. 본 연구에서는 조종 낙하산 하방에 탑재된 카메라가 헤딩 방향과 정렬되어 수직 하방을 바라보는 영상을 획득할 수 있으며, 탑재 데이터베이스를 활용하거나 별도의 처리 기술을 통해 영상 내에서 착륙을 지양해야 하는 장애물에 대한 정보를 기획했다고 가정한다. 이를 특히 식 (4)와 같이 중심점, 단/장반경, 장반경이 북쪽을 기준으로 회전한 방향(Orientation)으로 표현한 다양한 타원들의 집합으로 나타낼 수 있다고 가정한다.

$$H = \bigcup_{i=1}^{N_h} \left\{ \begin{array}{l} \left(x + \mu_x^{[i]}, y + \mu_y^{[i]} \right) \\ \left[\begin{array}{l} x \\ y \end{array} \right] = \left[\begin{array}{cc} c_{\theta^{[i]}} & -s_{\theta^{[i]}} \\ s_{\theta^{[i]}} & c_{\theta^{[i]}} \end{array} \right] \left[\begin{array}{l} x' \\ y' \end{array} \right], \\ \frac{x'^2}{a^{[i]^2}} + \frac{y'^2}{b^{[i]^2}} \leq 1 \end{array} \right\} \quad (4)$$

여기서 c, s 은 상응하는 삼각함수의 줄임, $\mu_x^{[i]}, \mu_y^{[i]}$ 는 i 번째 타원의 중심점, $\theta^{[i]}$, $a^{[i]}$, $b^{[i]}$ 는 i 번째 타원이 회전한 정도와 장반경, 단반경을 나타내고 평면 내 타원 밀도를 고려하여 전체 타원의 개수 $N_h \in N$ 를 설정한다. 이때, k 번째 시간에서 카메라의 FOV에 포함된 장애물들을 캐노피 헤딩 방향을 고려해 사영하여 점유 그리드로 표현한 I_k 의 (u, v) 위치에서의 값은 식 (5)와 같다. ($u, v \in N, u \leq U, v \leq V$)

$$I_{k,(u,v)} = \begin{cases} 1 & \text{if } \left(\mathcal{C}_{\psi_{p,k}} \left(\left[\begin{array}{l} u \\ v \end{array} \right] - \left[\begin{array}{l} U/2 \\ V/2 \end{array} \right] \right) \right) \frac{|X_{k,3}|}{f} + X_{k,1:2} \in H \\ 0 & \text{else} \end{cases} \quad (5)$$

여기서 f 는 초점 거리, X_k 는 k 번째 시간에서의 조종 낙하산 위치($=X_c$), $\mathcal{C}_{\psi_{p,k}}$ 는 ψ_p 만큼의 z 축 회전을 유도하는 2차원 회전 행렬, U, V 는 그리드의 크기이다.

본 강화학습 게임의 관측값은 최근 n 개의 스텝에서 획득한 점유 그리드를 적층한 형태로 활용하여, 시간 변화와 관련된 상태변수인 비행속도, 하강률 등을 간접적으로 추론할 수 있도록 식 (6)과 같이 설계한다. 해당 식에서 $I_k = I_{k,(::)}$ 를 만족한다.

$$x_{RL} = I_{(k-n+1):k}, (x_{RL(1)} = I_{k-n+1}, x_{RL(-1)} = I_k) \quad (6)$$

텐서 형태의 적층된 점유 그리드를 강화학습의 관측값으로 사용하므로 입력단에서 n 개의 채널을 가지는 CNN 구조에 완전연결(Fully Connected) 층을 추가해 평탄화한 Actor를 사용한다. 보상함수는 사전 지정한 위치 범위를 벗어나거나 캐노피가 착륙 시점에 대해 식 (7)과 식 (8)과 같이 설계한다.

$$r_{\text{glide}} = -\tan \left(\frac{\sum_{i=1}^U \sum_{j=1}^V P_{(i,j)} x_{RL(-1)(i,j)} \left/ \sum_{i=1}^U \sum_{j=1}^V P_{(i,j)} \right. \right) \quad (7)$$

$$r_{\text{land}} = \begin{cases} -10 & \text{On obstacle} \\ 10/\bar{v}_{td} & \text{else } \bar{v}_{td} > 1 \\ 10 & \text{else} \end{cases} \quad (8)$$

위험 지역으로 착륙하지 않는 것을 최우선으로 고려하였기 때문에 해당 지역으로의 착륙이 기대되는 경우 음의 보상을 주어 회피하도록 유도하는 부정 강화의 형태로 설계한다. 식 (7)에서 P 는 공간적으로 가중된 패치(Patch)를 의미하며, 그리드의 중심으로 갈수록 커지는 피라미드형 가중치를 가진다. 이는 관측값 내에서 가운데에 있는 장애물에 대해 큰 음의 보상을 인가하기 위함이다. $\tan(x)$ 함수는 $\tan(\pi/2(x-1/2))$ 의 줄임으로 $[0,1]$ 의 범위로 정규화된 입력을 $[-\pi/2, \pi/2]$ 로 대응시킨 후 일반 \tan 함수를 적용한 것이다. \bar{v}_{td} 는 착륙 시의 속력 $|v_c|$ 를 사전 설정한 연착륙의 판단 기준 속력으로 나누어 정규화한 값이다. 학습은 SAC⁽⁵⁾ 알고리즘을 사용하여 학습하였으며, 모의실험을 통해 획득한 장애물 위에서의 조종 낙하산 착륙 궤적을 Fig. 2에 나타내었다. 학습된 정책은 지속적으로 장애물과 멀어지려는 조종 명령을 생성하며, 열린 공간에서는 나선형 궤적을 통해 착륙을 시도하는 경향을 보인다.

본 연구에서는 상태 모델의 부재한 상황에서 적합한 게임 설계 및 심층 강화학습을 통해 조종 낙하산의 지상 장애물 회피 착륙을 유도하였다.

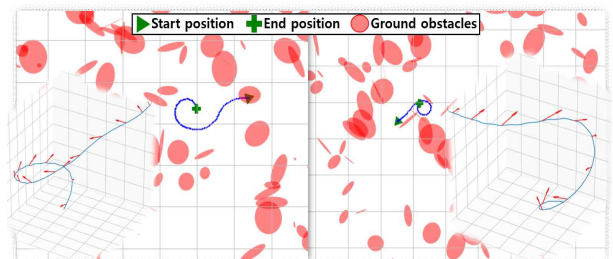


Fig. 2. Sample trajectories(2D & 3D) of a parafoil landing over obstacles

참고문헌

- 1) Taylor, A. P., Machin, R., Royall, P., and

Sinclair, R., "Developing the Parachute System for NASA's Orion: An Overview at Inception," In 19th AIAA Aerodynamic Decelerator Systems Technology Conference and Seminar, 2007, p. 2577.

2) Blackmore, L., "Autonomous precision landing of space rockets," *In Frontiers of Engineering: Reports on Leading-Edge Engineering* from the 2016 Symposium, Vol. 46, 2016, pp. 15-20.

3) Prakash, O., and Narayan A., "Modeling and simulation of 9-DOF parafoil-payload system flight dynamics," *AIAA Atmospheric Flight Mechanics Conference and Exhibit*, 2006, p. 6130

4) Lissaman, P., and Brown, G., "Apparent mass effects on parafoil dynamics," *Aerospace Design Conference*, 1993, p. 1236.

5) Haarnoja, T., Zhou, A., Abbeel, P., and Levine, S., "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," *International Conference on Machine Learning*, 2018, pp.