디지털산업정보학회 논문지 제20권 제3호 - 2024년 9월
http://dx.doi.org/10.17662/ksdim.2024.20.3.071

드론 군집 비행을 위한 다중 에이전트 최신 기술 분석 및 통신 최적화 기술 연구*

김 은 수** · 장 연 주** · 방 정 호***

Recent Trends in Multi-Agent Technology and Communication Optimization Research for Swarm Flight of Drones

Kim Eunsu · Jang Yeonju · Bang Jongho

〈Abstract〉

Artificial intelligence can be cited as a key linkage technology for expanding drones' application fields, and drones combined with artificial intelligence are expected to improve drones' operational capabilities based on algorithms that can solve complex tasks through learning. The purpose of this study is to analyze various latest research cases that apply deep reinforcement learning to drones to solve limitations for performing swarm flight and to propose a new research direction that applies them to multi-agent communication optimization technology. The process of the research is to investigate and analyze the methods for efficient operation of control and communication technologies required for swarm flight to be successful, and to apply algorithms that have the advantage of exchanging richer feedback between agents and having less learning than conventional methods when learning deep reinforcement learning algorithms. It is expected that the efficiency and performance of learning communication protocols optimized for swarm flight will be improved, which will increase the efficiency of mission performance when exploring or scouting large areas through swarm flight in the future.

Key Words : Swarm Flight, Multi-Agent Reinforcement Learning, Drone Control System, Drone Communication Protocols, DIAL

Ⅰ. 서론

드론 산업이 점점 성장함에 따라 드론에 관한 연구가 활발히 진행되고 있으며, 새로운 드론 기술들이 제시되고 있다. 드론은 토지 측량이나 군사적 정찰, 재난 상황 파악 및 인명 구조와 같은 분야에서 주로 활용된다. 그중 여러 대의 드론이 협력을 통해 임무를 수행하는 군집 비행에 대한 연구가 활발히 진행 중이다[1].

군집 비행은 다수의 드론에 의해 운영되므로 일부 드론에 고장이 발생해도 다른 다수의 드론들이 임무를 수

* 이 논문은 서울여자대학교 학술연구비의 지원에 의한 것임 (2024-0139)
** 서울여자대학교 소프트웨어융합학과 학부생
*** 서울여자대학교 소프트웨어융합학과 교수(교신저자)

행할 수 있기 때문에 그 연속성을 유지할 수 있다. 또한 넓은 지역을 탐사 또는 정찰할 때 여러 대의 드론을 통해 진행하므로 임무 수행의 효율성 즉, 넓은 면적에서 수행 임무가 주어질 때 광범위한 정보 수집 가능성을 높일 수 있어 경제적인 측면에서 효율성이 뛰어나기에 그 필요성이 더욱 부각된다[2]. 또한 군집 비행은 목표물 탐지 및 추적 시 다각도에서 인식한 객체에 대한 정보에 따라 추적의 정확도를 향상 시킬 수 있으며, 화물 운반 시에는 다량의 물품을 한 번의 비행으로 운반할 수 있는 등, 다양한 측면에서 그 효과를 보인다.

군집 비행이 성공적으로 수행되기 위해서는 안정적인 시스템, 제어 기술 및 통신 기술이 요구된다. 군집 비행은 여러 대의 드론이 편대를 이루어 움직이기 때문에 드론간 충돌 발생 확률이 높으며, 실외에서 진행되는 군집 비행 특성상 편대 내에서 변수가 발생하기도 한다. 따라서, 드론의 군집 비행에서는 이를 방지하기 위한 안정적인 시스템과 드론의 복잡한 편대 제어를 위한 제어 기술이 필요하다. 뿐만 아니라, 드론의 제어 명령을 포함한 위치 정보를 주고받기 위하여 지상국 컴퓨터 시스템과 수많은 드론간의 통신 그리고 편대 내 드론간의 통신 등이 요구되므로 이를 효율적으로 지원할 수 있는 통신 기술도 필수적이라 할 수 있다[3].

드론의 군집 비행과 같이 여러 개의 비행 물체가 서로 협력하여 임무를 수행하고 발생한 문제를 해결하기 위해서는 대표적인 인공지능 기술인 강화학습 기술이 적합하다. 이에 따라 최근 연구에서는 군집 드론의 자율 비행을 성공적으로 수행시키기 위하여 드론의 비행에 강화학습을 접목시키는 사례들이 늘고 있다. 강화학습이란 에이전트가 환경과 상호작용하며 특정 행동에 대한 보상을 받아 목표를 달성하는 방법을 학습하는 인공지능의 한 분야이다. 이를 드론 비행에 적용 시키면 드론이 에이전트가 되고, 드론이 비행하는 공간이나 지형 또는 장애물들이 환경이 되고, 자율 비행하는 과정에서 취할 수 있는 모든 가능한 동작들이 행동이 된다. 이때, 드론이 최적의 과정으로 임무를 수행할 수 있도록 반복적으로 학습하는 것이 바로 드론 자율 비행에 강화학습을 적용한 것이다[4].

동적 장애물에 대한 위험과 비행 환경의 불안정성이 드론 비행의 본질적인 특성임을 고려할 때, 강화학습이 적용되지 않은 기존의 자율 비행은 이러한 불확실성에 대응하기 어렵다는 한계가 있다. 그 결과, 강화학습을 접목시킨 군집 비행 기술이 점점 더 필수적으로 인식되고 있다. 그러나, 드론이나 로보틱스 분야에서는 그 상태가 다차원 규모로 계속해서 변화하므로 그 특성상 고전적인 강화학습을 적용하는데 제약이 존재한다. 따라서, 최근에는 기존의 강화학습에 딥러닝을 적용시킨 심층 강화학습 분야가 연구되고 있는 추세이다. 그중에서도 드론이나 로보틱스 분야의 특성을 고려해 여러 에이전트가 협력 또는 경쟁을 통하여 높은 보상을 얻고자 학습하는 방법에 대한 연구가 이루어지고 있는데, 이를 다중 에이전트 강화학습이라 한다[5]. 드론의 군집 비행을 위해서는 단일 에이전트가 아닌 다른 에이전트와의 관계, 다른 에이전트로부터 오는 영향을 함께 고려하는 다중 에이전트 간의 협력 기술이 필수적으로 요구된다. 심층 강화학습을 적용시킨 드론의 군집 비행이 성공적으로 이루어지기 위해서는 제어 기술의 발전이 필수적이며, 제어 기술은 안정화 제어와 경로 계획 두 가지 측면에서 연구되고 있다. 안정화 제어 연구는 드론의 충돌 방지와 안정적인 비행을 목표로 하며, 경로 계획 연구는 드론이 목적지까지 안전하게 이동할 수 있는 최적의 경로를 탐색하는 것이다. 이러한 심층 강화학습을 적용한 드론의 군집 비행은 기존의 자율 비행 한계를 극복하고 부가적인 기대효과를 제공함을 연구를 통해 확인할 수 있었다[6].

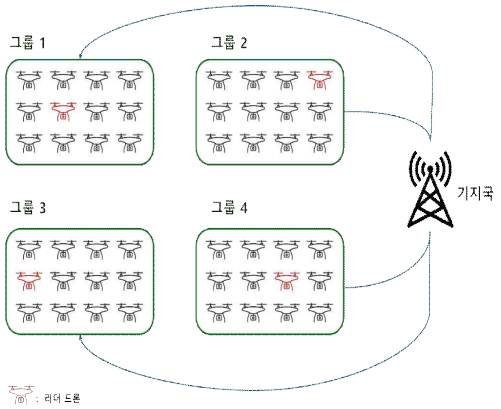
따라서, 본 연구에서는 드론이 군집 비행을 수행하는 과정에서 발생하는 한계점과 이러한 한계점을 해결하기 위해 드론에 심층강화학습을 적용한 다양한 최신 연구 동향을 소개하고 분석한다. 또한 효율적인 군집 비행을 위한 제어 구조 및 드론 센서값을 포함한 통신 메시지 포맷을 정의하고, 이를 다중 에이전트 통신 최적화 기술에 적용시킨 새로운 연구 방향을 제안한다.

Ⅱ. 관련 연구

2.1 제어기술

Bang et al.[7]에서는 군집 드론의 제어를 위한 시스템 구조를 그림 1과 같이 제안하였다. 시스템은 크게 기지국, 리더 드론, 리더 드론을 제외한 나머지 드론으로 구성된다. 수십 대의 드론이 군집 비행을 위해 여러 개의 드론 그룹으로 구성되었고, 기지국은 각 드론과 통신을 한다. 또한 드론 그룹에 변동사항이 생길 경우를 위해 통신 우선 순위를 위한 프로토콜이 필요하다. 그리고 이러한 프로토콜은 드론의 비행 과정에서 생길 수 있는 통신 오류에 대비할 수 있도록 설계되었다. 또한 그룹 내 드론들에는 통신 순위가 결정되어 있다. 드론이 자체적으로 순위를 정할 수 있어야 하며 통신 순위가 최우선이 된 드론은 드론 그룹의 리더 드론이 된다. 리더 드론이 그룹에서 빠질 경우 다음 순위의 드론이 최우선 드론으로 교체될 수 있도록 알고리즘이 설계되어 있다.

기지국의 역할은 우선 사용자로부터 입력을 받으며 드론 그룹의 구성을 관리한다. 예를 들어, 드론 그룹에 새로운 드론이 들어오면 적절한 위치에 할당해야 한다. 또 드론 그룹에 비행 경로를 제시해주며, 비행 상태를 확인하고 문제점을 파악할 수 있어야 한다. 리더 드론은 기지국과 통신하며 각 드론들의 정보를 수집하고, 수집한 정보에 따라 비행 경로를 전달한다. 그리고 드론에 이상이 생겼을 경우에 이를 기지국에 전달해주고 대책을 마련한다. 각 그룹의 드론들은 위치 보정 장치를 통하여 비행을 수행하고 자신의 정보를 리더 드론에 전달한다. 실내와 실외 군집 비행에서의 제어 기술을 고려한 연구에서는 우선 실내 군집 비행 제어 방법으로 PID 제어기와, 드론의 위치 정보 인식을 받기 위해 모션 캡쳐 기술을 사용하였다. 모션 캡쳐 기법을 통해 지상국 시스템은 위치 정보를 전달받을 수 있으며, 위치 정보를 통해



〈그림 1〉 드론 군집 비행에 대한 시스템 구조

현재 위치($P_{current}$)와 목표 위치(P_{target})간 차이를 식(1)을 통해 3차원 벡터 형태로 구현할 수 있다. 식(1)에서 구한 오차값은 절대 좌표계의 값이기 때문에 Direction Cosine Matrix (DCM)을 통해 동체 좌표계로 변경해준다[8].

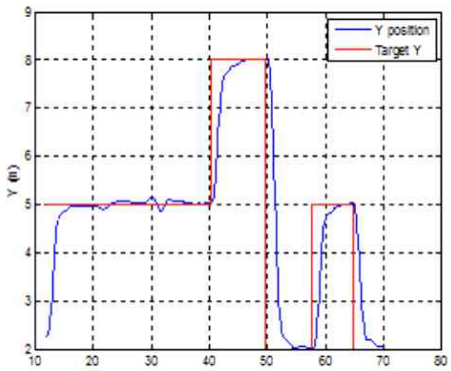
$$e(t) = P_{target} - P_{current} \quad (1)$$

변경된 오차값은 PID 제어기를 통해 Roll/Pitch/Yaw 제어값인 C(t)를 구할 수 있다. C(t)를 구하는 식은 식(2)와 같다. K_p는 비례항 이득값 (gain value)을 나타내고, K_i는 적분항 이득값을 나타낸다. 적분항을 구할 때 20cm 안으로 들어오는 경우에 한해 오차값을 누적하였는데 이때 v는 20cm안에 들어오는 시간을 의미한다.20cm를 벗어나는 경우 적분항은 0이 된다. K_d는 미분항 이득값인데 이는 오차와 응답속도 보완을 위해 사용된다. 여기서 K_p값이 증가하면 출력값이 목표값에 빠르게 도달하지만 너무 크면 오버슛이 일어나게 되고, K_p값이 감소하면 출력값이 목표값에 느리게 도달하며 제어 값이 작아져 정상상태 오차가 커지게 된다. K_i는 비례 제어를 사용했을 때 나타나는 정상상태 오차를 해

결하기 위해 고안되었다. 정상상태 오차가 나타나면 오차를 계속 적분하는 방식으로 오차를 줄여준다[8].

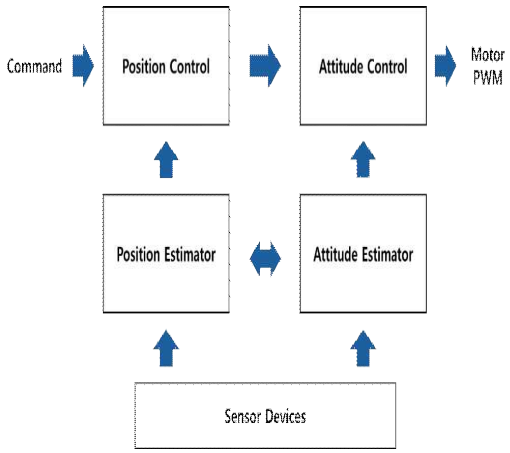
$$C(t) = K_p e(t) + K_i \int_v^t e(t) + K_d \frac{d}{dt} e(t) \quad (2)$$

식(1)과 식(2)를 수행한 결과인 그림 2를 통해 목푯값과 실제 Y 값이 얼마나 일치하는지 알 수가 있다. 여기서 목푯값은 드론이 도달하고자 하는 목표 위치, 실제 Y값은 현재 드론의 위치를 의미한다. 드론이 비행하면서 실시간으로 자신의 위치를 계속 갱신하고, 이 값을 바탕으로 목푯값과의 차이를 계산한다. 그림 2에서는 목푯값과 실제 Y 값이 유사한데 이는 PID 제어기를 통해 드론이 목표 위치에 가까이 도달했음을 의미한다.



〈그림 2〉 PID 제어기를 통한 Y축의 위치제어 특성

실외 군집 비행의 제어 방법으로는 PID제어기를 사용했다는 점이 실내와 유사하다. 하지만 실외 비행의 경우 제어 방법을 드론의 기체 내에서 수행해야 하기 때문에 위치 제어 기능을 드론 내부에 가지고 있어야 한다. 따라서 실외 비행의 제어 방법으로 그림 3과 같은 구조를 제안하였으며, RTK-GPS 센서 융합을 통한 위치 예측을 위해 위치 추정기(Position Estimator)를 변경하고, 위치 제어기(Position Control)의 동작 방식을 수정하였다[8].



〈그림 3〉 실내외 군집 비행 시스템 구조

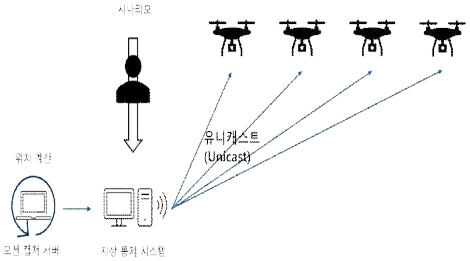
제어 기술에 강화학습을 적용한 연구에서는 자율 비행 및 회피를 수행하기 위해 강화학습을 적용한 드론의 제어 및 자세 안정화 기법을 제안하였다[9]. 이 기법은 비행체의 데이터를 기반으로 수학적으로 제시하기 어려운 모델을 간접적으로 찾아내, 이를 바탕으로 직접 구현하는 방식이며, 경로를 생성한 후 강화학습을 활용한 알고리즘을 독립적으로 설계했다.

2.2 통신기술

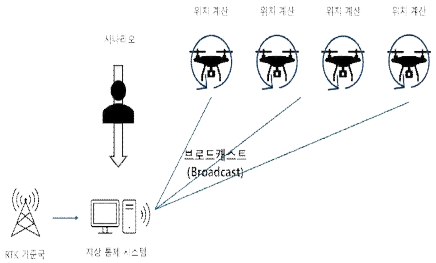
실내 군집 비행을 위해 주로 사용하는 통신 방법은 WiFi이며, 그림 4와 같은 구조를 가지며 유니캐스트 방식으로 명령을 전달한다. 하지만 WiFi를 사용하는 경우 통신이 불규칙적으로 끊기게 되는데 이러한 경우 실시간성을 보장받지 못하는 문제점이 생긴다. 따라서 불필요한 데이터를 최소화하고, 무선 네트워크의 채널 수를 늘리는 방법으로 문제를 해결하고자 했다.

실외 군집비행에서의 통신도 실내 군집비행과 마찬가지로 WiFi를 사용한다. 실외의 경우는 그림 5와 같은 구조를 가진다. 실외 군집 비행은 보다 넓고, 다양한 환경을 접하므로 실내 방식과 비교했을 때 통신 방법과 통신

양이 다르다. 드론 내에서 자체적으로 위치 정보를 관리하므로 보정 신호만 전달해 주면 된다. 지상국은 브로드캐스트 방식으로 보정 신호를 드론에게 전달하는데, 브로드캐스트 방식을 사용하면 드론들에게 보정 신호를 한 번에 전달할 수 있어서 통신량이 감소된다. 또한 명령 정보를 드론 내부에 삽입하여 GPS신호를 동기화 시키므로 명령을 전달할 필요가 없어진다. 사용자는 작성한 시나리오 파일을 드론들에게 지상국을 통해 전달하고, 드론 내부에 저장해 이에 맞춰 비행을 수행한다[8].



〈그림 4〉 실내 군집 비행 통신 시스템



〈그림 5〉 실외 군집 비행 통신 시스템

2.3 다중 에이전트 강화학습 알고리즘

1장에서 소개한 것처럼 성공적인 군집비행을 위해 중요한 요소들이 여러 가지 있지만, 그중에서도 안정적인 시스템을 구축하는 것이 핵심이다. 최근 관련 연구에서는 안정적인 시스템 구축을 위해 다양한 다중 에이전트 강화학습 알고리즘을 사용한다.

2.3.1 강화학습 및 심층 강화학습 알고리즘

강화학습 알고리즘은 가치 함수와 정책 중 무엇을 최적화하는지에 따라 가치 기반 방식(Value-based Method)과 정책 기반 방식(Policy-based Method)으로 구분된다. 가치는 어떤 상태에서 얻을 수 있는 반환값의 기댓값을 의미하며, 이때 반환값이 어떤 상태에서의 총 보상이면 상태-가치 함수라 하고, 어떤 상태에서 특정 행동을 취했을 때의 총 보상이면 행동-가치 수함라고 한다.

그리고 가치 함수 기반 알고리즘(Value-based Algorithms)은 가치 함수를 이용해 보상을 최대화하는 행동을 반복적으로 학습하여 최적화하는 알고리즘으로, Q-learning, DQN(Deep Q-Network) 이 대표적이다.

Q-learning과 DQN은 서로 밀접하게 관련되어 있는데, Q-learning은 가장 기초적인 model-free 강화학습 알고리즘으로 특정 상태에서 특정 행동을 취할 때의 예상 보상을 나타내는 Q-값(Q-values)을 업데이트 하면서 최적의 정책을 학습한다[7]. 이때, Q-learning은 각 상태와 행동 세트(s,a)의 Q-값을 테이블 형태로 만들고, 이 테이블을 통해 상태와 행동의 모든 경우에 대해 Q-값을 업데이트 한다.

Q-learning 알고리즘이 Q-값을 업데이트 하는 과정에서 사용되는 행동 가치 함수는 식 (3)의 Bellman Equation을 기반으로 각 sequence 마다 업데이트 된다. Bellman Equation은 어떤 상태에서 특정 행동(s,a)을 취했을 때 받을 수 있는 보상의 모든 기대값을 의미하고, 그 보상의 기대값을 최대화 할 수 있는 다음 행동(a')를 선택하는 매커니즘으로 강화학습에 적용된다.

그러나, 이러한 메커니즘의 Q-learning은 일반화가 어렵고 '차원의 저주'라고 불리는 문제 때문에 상태와 행동의 수가 많은 경우에서는 적용의 한계가 있다. 특히, 로보틱스나 드론 분야는 상태와 행동이 연속적일 뿐만 아니라 수많은 변수로 구성된 고차원 상태공간이므로 더욱 그렇다.

$$Q^*(s,a) = E_{s' \sim \epsilon}[r + \gamma \max_{s'} Q^*(s',a')|s,a] \quad (3)$$

이러한 한계를 극복하기 위해 제안된 것이 바로 DQN이다. DQN은 기본적으로 Q-learning의 개념을 취하지만, 심층 신경망(Deep Neural Network)을 통해 Q-값을 근사한다는 점에서 차이가 있다. DQN은 딥러닝 네트워크를 사용하므로 합성곱 신경망(CNN)을 사용하여 이미지를 처리하며, 경험 재현(Experience Replay)을 채택하여 성능을 개선한 알고리즘이다[3].

반면, 정책 기반 알고리즘(Policy-based Algorithms)은 가치 함수 없이 에이전트가 정책을 직접 학습하여 최적화하는 알고리즘으로, 이때 정책은 특정 상태에서 보상을 최대화할 수 있는 행동을 선택하는 방법을 의미한다. 정책 기반 알고리즘의 대표적인 예시로는 몬테카를로 방법(Monte-Carlo Method), DDPG 알고리즘이 있다.

심층 강화학습의 정책 기반 알고리즘에 대한 연구는 대부분의 알고리즘들이 Actor-Critic을 기반으로 이루어져 있다. 심층 강화학습에서는 딥러닝의 개념을 가져오므로 정책이 신경망 형태로 표현되기도 하는데, 바로 Actor가 입력값을 가지고 에이전트의 행동을 출력하는 정책 신경망이다.

가치 함수 기반 알고리즘은 특정 상태에서의 어떤 행동의 가치를 계산하고, 정책 기반 알고리즘은 특정 상태에서 취할 수 있는 어떤 행동 자체를 계산해 최대의 보상을 얻을 수 있도록 최적화하는 것이다.

강화학습 알고리즘은 환경에 대한 모델의 유무에 따라 Model-based와 Model-free 알고리즘으로 나뉘는데, Model-based 알고리즘은 학습을 진행하면서 얻어진 결과인 보상을 통해 최적화하는 것이 아닌 모델의 동작 방식을 학습해 보상을 예측하는 방식으로 최적화하는 알고리즘을 말한다. 반면, Model-free 알고리즘은 환경과 직접 상호작용하며 각 상태에 따른 행동을 선택하고 그 선택에 따른 보상을 학습하면서 최적화하는 알고리즘이다.

이렇게 강화학습에는 서로 다른 다양한 알고리즘이 존재하므로, 해결하고자 하는 문제의 특성에 따라 최적의 알고리즘을 선택해 학습을 진행할 수 있다. 그러나, 본 연구에서 다루고자 하는 드론의 군집 비행을 위한 인공지능 솔루션을 얻고자 하는 경우에는 일반적인 강화학습에서 사용되는 알고리즘을 적용하는 부분에 있어서 한계가 있다. 드론 비행은 상태 공간에 연속적인 변화와 변수가 존재하며, 그 중에서도 군집 드론은 다중 에이전트로 구성된다는 환경의 특수성이 있기 때문이다. 이때, 이러한 특수성을 고려해 군집 드론을 위한 인공지능 연구는 다양한 방식으로 진행되고 있는데, 그 사례는 다음과 같다.

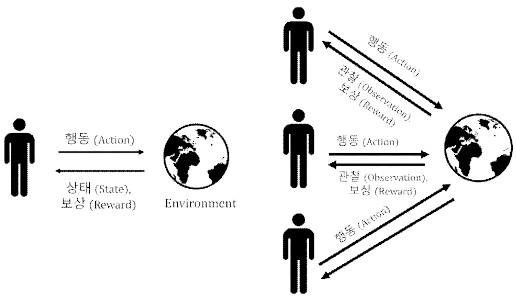
드론의 건축자재 인식을 위해 딥러닝 모델 중 하나인 Yolo를 사용하는 연구가 있다. Yolo는 R-CNN 기반의 한 모델인데 자재의 윗면과 측면이 나온 영상으로 학습시킨다. 단 자재 더미 영역을 관심 영역으로 설정한 후 영상 크기를 448x448로 지정한다. 또한 학습에 사용할 자재로는 벽돌, 시멘트, 연석 세 가지만 사용하며, 학습은 검증 시 발생하는 평균 오류율이 1.0 이하가 될 때 종료한다[10].

강화학습을 기반으로 한 다중 객체 추적 시스템을 제안하는 연구도 있다. 다중 객체 추척 시스템은 두가지 방법을 수정 후, 조합하여 적용하였다. 첫 번째로 SORT를 기반으로 하며, 두 번째로는 Re3방법을 기반으로 한다[11, 12].

2.3.2 다중 에이전트 강화학습 협력기술

그림 6을 통해 확인할 수 있는 것처럼 다중 에이전트 강화학습은 환경과 상호 작용할 뿐만 아니라 그룹을 구성하는 다수의 에이전트들끼리 서로 상호작용 한다는 점에서 단일 에이전트 강화학습과 차이가 있다. 군집 드론의 경우 다중 드론이 각기 다른 관측을 갖는 동시에 통합된 행동을 통해 군집 비행의 목표를 달성해야 하는데 이러한 상황에서 사용하는 것이 바로 다중 에이전트 협력 기술이다. 이는 제한된 상황에서 다중 에이전트 간의 협력을 위해 모색한 기술로 중앙형 학습과 분산형 실행(Centralized Training Decentralized Execution, CTDE)을 통해 구현된다[13].

예를 들어, 드론이 실제 분산되어 비행할 때는 다른 드론의 관측 각도나 고도, 배터리 상태 등 관측 정보를 파악할 수 없지만 학습 단계에서는 이를 가능하도록 하여 더 다양한 경우에 대해 학습을 진행하고 이를 통해 드론 간의 협력을 보다 성공적으로 달성할 수 있는 환경을 말한다. 이때, 중앙형 학습과 분산형 실행을 구현하는데 가장 대표적으로 사용되는 다중 에이전트 강화학습 알고리즘은 COMA와 MADDPG이다.



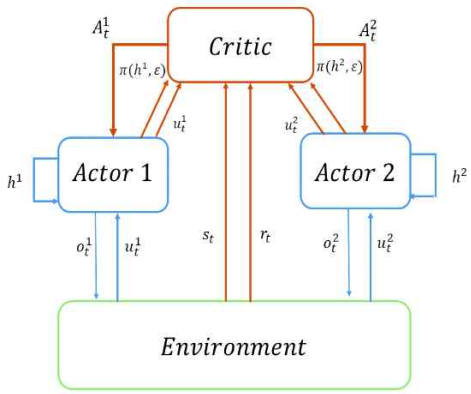
〈그림 6〉 단일 에이전트 강화학습과
다중 에이전트 강화학습의 기본 구조

(1) Counterfactual Multi-Agent (COMA)

COMA는 CTDE의 가장 기본적인 알고리즘 중 하나로, 그림 7을 보면 Centralized Critic(중앙형 크리틱)을 가짐과 동시에 모든 에이전트가 오직 하나의 Critic을 사용하므로 중앙화된 크리틱이 전체 에이전트의 모든 행동에 대한 가치를 평가하고, 이를 통해 전체 시스템의 성과를 평가하고 보상을 할당한다. 이러한 특징에 의해 COMA 알고리즘은 Credit Assignment 문제 상황에서 주로 사용된다.

Credit Assignment 문제란 다중 에이전트 환경에서 각각의 에이전트의 행동에 대해 보상을 할당하는 과정에서 생기는 문제로, 한 에이전트의 특정 행동이 전체적인 성과에 어떤 영향을 미치는 지를 정확하게 파악하는 것이 어렵기 때문에 이를 해결하기 위해서 Centralized Critic을 사용하는 것이 효과적이다[15].

그러나, 이렇게 모든 에이전트의 행동에 대한 가치를 평가하는 COMA 알고리즘의 특성상 Credit Assignment 문제 해결에 강하고 간단한 문제에서는 비교적 높은 정확도를 보일 수 있지만 다중 에이전트 환경에서는 에이전트 수가 증가 할수록 연산량에 대한 부담이 비례하기 때문에 한계를 갖는다.



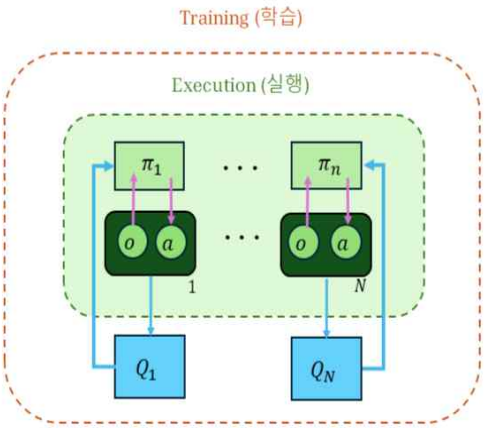
〈그림 7〉 COMA 알고리즘의 구조

(2) Multi-Agent Deep Deterministic Policy Gradient (MADDPG)

MADDPG는 Centralized Critic(중앙형 크리틱) 구조라는 점에서 COMA 알고리즘과 유사하지만, 그림 8을 보면 모든 에이전트가 개별적인 Critic을 가지므로 각 에이전트가 자신의 정책을 개별적으로 업데이트 하여 학습시키는 것을 목적으로 한다. 또한 대표적인 정책 경사 기반 DDPG 알고리즘을 다중 에이전트 환경으로 확장시킨 것으로 다른 에이전트의 영향을 고려하기 위해 상태와 행동을 직접 행동-가치 함수(Action-Value Function)의 입력으로 사용하여 정책을 찾는다. 행동-가치 함수는 특

정 상태에서 어떤 행동의 가치 즉, 총 보상의 기댓값을 계산해 다른 에이전트의 영향을 직접 고려하기 때문에 협력 또는 경쟁, 혼합 협력-경쟁 시나리오 모두 적용이 가능하다.

MADDPG는 DDPG와 비교 했을 때, 다중 에이전트 환경에서 더 높은 성능의 정책을 찾을 수 있지만 모든 에이전트의 상태와 행동을 입력으로 사용하기 때문에 에이전트의 수가 증가하면 에이전트의 상태와 행동의 수도 비례하게 되므로 정책 경사(Policy Gradient)를 구하기 위한 목적함수가 상당히 복잡해진다는 단점이 있다[16].



〈그림 8〉 MADDPG 알고리즘의 구조

(3) Mean Field Reinforcement Learning (MF-RL)

MF-RL 알고리즘은 다중 에이전트 환경에서 에이전트 간의 상호작용을 고려하기 위해 개발된 알고리즘으로, 내쉬 균형(Nash Equilibrium) 전략을 어떻게 찾는 지에 따라 MADDPG와 서로 구분된다. 내쉬 균형은 게임 이론에서 사용되는 개념으로, 군집 비행에 적용하면 모든 에이전트 즉, 드론이 자신에게 최적인 전략을 선택한 상태를 의미한다. 이때, MF-RL 알고리즘은 식(4)의 Mean Field Approximation(평균장 근사)를 이용해 각 에이전트가 다른 에이전트의 평균 행동을 기반으로 학습하여 상호작용의 복잡성을 줄이는 방식을 통해 내쉬 균형에 도달하며 평균 행동을 정책 경사의 목적함수의 입력으로 사용하거나 행동-가치 함수의 입력으로 사용하는 알고리즘이다.

$$a^k = \bar{a}^j + \delta a^{j,k}, \quad \text{where } \bar{a}^j = \frac{1}{N^j}\sum_k a^k \tag{4}$$

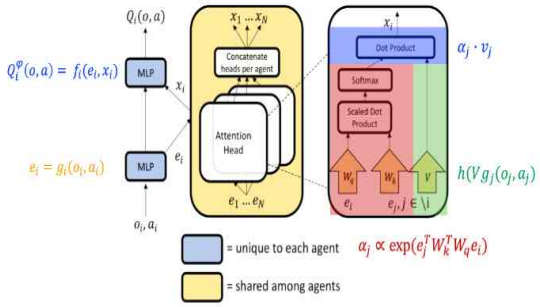
다른 에이전트의 영향을 주변 에이전트의 평균 행동이라는 하나의 인자로 압축해서 표현하기 때문에, 에이전트의 수가 많더라도 에이전트 간의 영향이 넓지 않은 편이며, 특히 행동이 복잡하지 않은 환경에서는 우수한 정책 또는 행동-가치 함수를 찾을 수 있다. 하지만 에이전트 간의 영향의 범위가 넓은 시나리오에서는 적용에 한계가 있다는 단점이 있다[17].

(4) Multi-Actor-Attention-Critic (MAAC)

MAAC 알고리즘은 인공 신경망 기반의 Attention 메커니즘을 사용하여 중앙화된 Critic을 학습하고 이를 통해 상황에 따라 에이전트 간의 중요도를 계산하며, 이를 정책 경사의 목적함수의 입력으로 사용하는 알고리즘이다. 어텐션 메커니즘은 딥러닝의 자연어 처리 기법 중 하나로 최근에는 다중 에이전트 강화학습 연구에서도 그 기법이 적용되고 있다. 기본적인 원리로는 쿼리, 키, 값으로 구성되어 쿼리와 키 간의 유사도를 계산하고 그 유사도 값을 이용해 가중치와 가중합을 산술적으로 계산함으로써 학습 모델이 중요한 정보를 선별할 수 있도록 한다. 뿐만 아니라, 싱글 에이전트 강화학습 알고리즘의 종류 중 하나인 Soft Actor-Critic 알고리즘을 기반으로 구성되어 협력 또는 경쟁 시나리오에서 모두 적용이 가능하다.

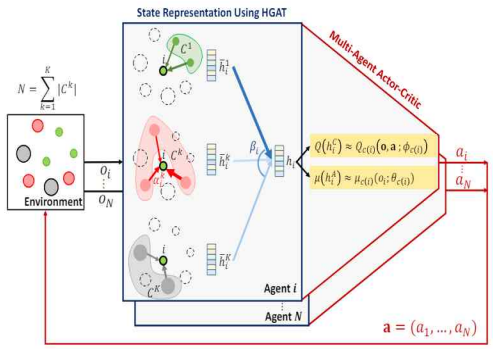
따라서 다른 에이전트의 상태와 행동을 직접 입력으로 사용하는 MADDPG 대비 에이전트 간의 관계를 더 효율적으로 모델링할 수 있고, 더 중요한 정보에 집중하므로

동일 환경에서 에이전트 수가 늘어날 때도 성능이 크게 저하되지 않으며 주어진 문제를 해결할 수 있다[18].



〈그림 9〉 Attention 적용된 Soft-Critic 기반의 COMA 알고리즘 모델 구조

이전트 정보를 통합 및 각 에이전트의 최종 상태를 업데이트 한다[19].



〈그림 10〉 HAMA 알고리즘의 구조

(5) Hierarchical graph Attention-based Multi-agent Actor-critic (HAMA)

HAMA 알고리즘은 어텐션 메커니즘을 적용한다는 점에서 MAAC과 유사하지만 구조와 특성에서 서로 구분된다. HAMA는 에이전트의 특성을 고려해 그룹을 만들고 이때 형성된 계층적 구조에 어텐션 메커니즘을 결합하여, 그룹별 어텐션을 계산한 뒤 이를 정책 경사에 적용한 알고리즘이다. 따라서, 에이전트로별 특성이 명확히 구분되고 그룹화가 용이한 환경에서는 HAMA 알고리즘이 기존 알고리즘보다 좋은 성능을 보인다. 또한 그룹별로 중요도를 계산하기 때문에 계층 내의 에이전트 수가 증가하더라도 성능 저하 없이 작동한다.

HAMA 알고리즘에서는 어텐션 메커니즘을 이용하여 강화학습을 수행할 때 목적에 맞게 네트워크를 적용하는데 그림 10은 HGAT(Heuristic Graph Attention Network)가 적용된 HAMA 알고리즘이다. HGAT은 세 가지 단계에 걸쳐서 각 에이전트 상태의 중요도를 업데이트 한다. 첫 번째로는 각 에이전트의 상태 정보를 임베딩 벡터로 변환하고, 두 번째로 임베딩된 벡터의 가중치를 계산하고, 마지막으로 계산된 가중치를 바탕으로 에

(6) 정리

다섯 가지의 다중 강화학습 알고리즘들은 CTDE 방식을 따른다는 점에서 공통적이나, 서로 다른 특징을 바탕으로 동작한다. 따라서 적용하고자 하는 시나리오나 에이전트 또는 모델의 사용 목적에 맞게 적절한 알고리즘을 선택하는 것이 중요하다. 다음은 이를 정리한 표이다.

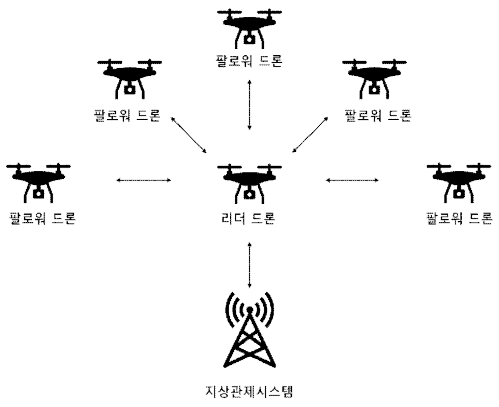
〈표 1〉 다중 강화학습 알고리즘의 비교

알고리즘	중앙형 크리틱	정책 업데이트 방식	확장성	시나리오
COMA	사용	중앙화	제한적	협력
MADDPG	사용	개별화	보통	협력 또는 경쟁
MF-RL	미사용	평균장 근사	좋음	단순 협력
MAAC	사용	어텐션 메커니즘	좋음	협력 또는 경쟁
HAMA	사용	계층적 어텐션 메커니즘	좋음	그룹 기반 협력 또는 경쟁

Ⅲ. 제안 방법

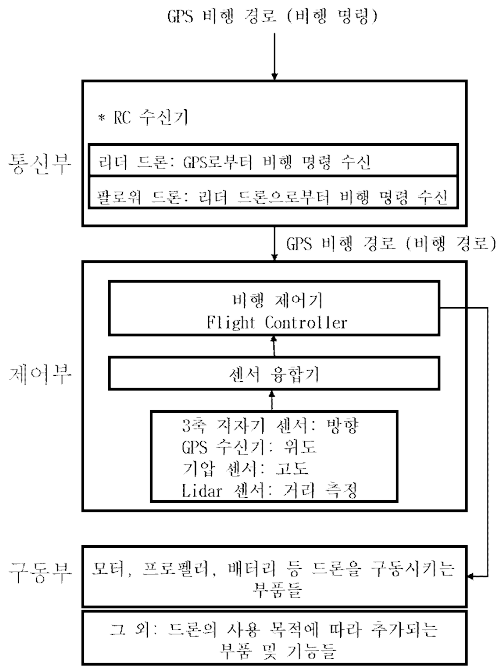
3.1 드론 군집비행 제어 구조

그림 11에서는 그룹에서 선택된 리더 드론은 지상관제시스템(Ground Control System, GCS)과 통신하며 전체 군집이 수행할 비행 임무 및 목적지에 대한 GPS 경로를 수신한다. 즉 모든 개별 드론은 리더를 통해 하나의 네트워크로 연결되는 구조를 갖는다. 이러한 방식으로 제어구조를 설정할 경우 GCS의 입장에선 리더 드론만 관리하면 군집을 이루는 나머지 모든 드론에 대한 제어가 가능해지므로 다방면에서 편리하다는 장점이 있으나, 리더 드론에게 결함 등의 문제가 발생할 경우 나머지 팔로워 드론들에 대한 제어를 한 순간에 잃을 수 있다는 위험성 또한 존재하므로 돌발 상황에 대한 대비가 필요하다. 리더 드론과 팔로워 드론들 간의 통신 방식으로는 높은 속도의 데이터 전송이 가능하며 하나의 채널만으로 제어 신호를 송수신할 수 있는 WiFi 무선 통신 프로토콜을 사용한다. 또한 리더 드론이 팔로워 드론에게 임무 및 목적지 GPS 경로를 송신할 때에는 브로드캐스팅 방식을 취하며, 각 팔로워 드론은 자신이 수신자로 지정된 메시지만 수신한다.



〈그림 11〉 리더-팔로워 군집 드론 제어 구조

3.2 개별 드론의 동작 구조 및 드론 군집비행 통신 메시지 포맷



〈그림 12〉 개별 드론의 동작 구조

그림 12는 군집을 이루는 개별 드론들의 동작 구조를 나타낸다. 통신부는 주로 RC 수신기 등의 단말기를 통해 비행 명령을 수신하는 역할을 수행하는데, 이때 리더 드론은 GCS로부터, 팔로워 드론들은 리더 드론으로부터 해당 명령을 전달받는다. 제어부는 통신부로부터 전달된 비행 명령을 수신하여 비행이 적절히 이루어지도록 제어 및 관리하는 역할을 담당하는데, 방향 정보를 수집하는 3축 지자기센서, 경도 및 위도 정보를 수집하는 GPS 센서, 그리고 각각 고도와 거리 정보를 측정하여 수집하는 기압 센서와 Lidar 센서 등 각종 센서 정보를 기반으로 센서 융합기를 거쳐 드론에서 CPU와 같은 두뇌의 역할을 담당하는 비행 제어기(Flight Controller)를 통해 처리되어 최종적인 제어 명령이 내려지게 된다. 마지막에 동

작하는 구동부는 모터와 프로펠러, 배터리 등 물리적인 장치들을 통해 드론을 구동시키는 역할을 담당한다. 일반적으로 드론의 비행 상태는 회전운동상태와 병진운동상태로 파악되며 각 상태는 자이로센서와 가속도센서를 통해 측정되지만 본 논문에서는 군집 드론의 비행에서 개별 드론들 간에 공유하여 통신과 협력의 효율을 높일 수 있을 것으로 보이는 주요 센서 네 가지를 중심으로 개별 드론의 동작 구조를 정의하였다[14][20].

군집을 이루어 비행하는 개별 드론들이 하나의 목표를 위해 협력하기 위해서는 각종 센서를 통해 파악되는 각자의 상태(state)를 정형화된 포맷에 맞추어 서로 간에 공유하는 것이 매우 중요하다. 다중 에이전트 통신에 최적화된 강화학습 기반 기술인 DIAL에서 군집을 이루는 에이전트들은 서로 간의 그라디언트(gradient)를 주고받으며 통신 프로토콜을 학습하게 되는데, 해당 기술은 광범위한 다중 에이전트 문제들에 적용될 수 있는 종합적인 개념이므로 보다 구체적인 문제인 군집 드론 비행 상황에 적용할 수 있는 통신 메시지 포맷이 필요하다.

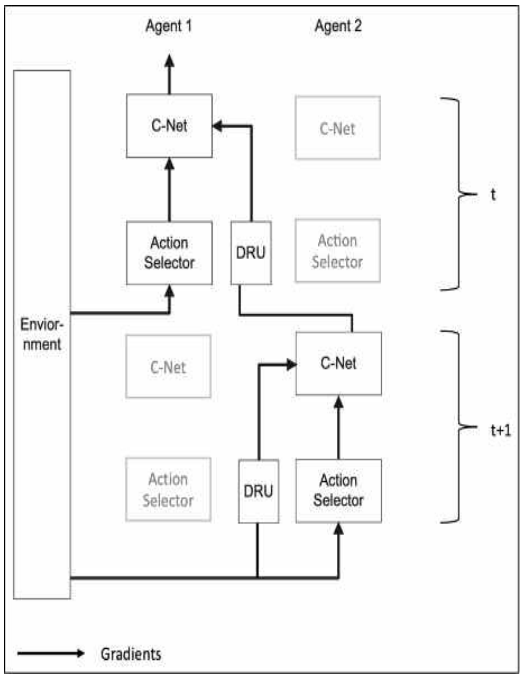
〈표 2〉 군집 드론 통신 데이터 포맷

No	Cell Name	의미
1	DroneNo	드론의 고유번호 (ID)
2	OrderNo	통신 순서
3	TargetNo	데이터 전송대상
4	GPSLoc1	GPS 센서 (X축)
5	GPSLoc2	GPS 센서 (Y축)
6	Barometer	기압 센서 (Z축)
7	Magnetometer	자력 센서 (3축 지자기 센서)
8	Lidar	라이다 센서
9	Command	명령어 코드
10	MSG	전달 메시지
11	CurTime	현재 시각
12	Reserve	예약 셀

표 2은 그림 9에 포함된 각종 센서를 중심으로 정의한 군집 드론에 적합한 통신 데이터 포맷을 나타낸다. DroneNo과 OrderNo, 그리고 TargetNo은 각각 드론의 고유 ID와 통신 순서, 데이터의 전송 대상을 의미한다. Cell No 4~8번이 드론의 센서값에 해당되는데, GPSLoc1과 GPSLoc2에는 GPS 센서 상 드론의 경도인 X축과 위도인 Y축 좌표가 입력되며, Barometer에는 기압 센서 상 드론의 고도인 Z축 좌표가 입력된다. Magnetometer는 3축 지자기 센서로 드론의 방향 정보를 담으며 Lidar 센서 값은 대기 중 물체에 대한 거리를 측정한 값을 나타내어 주변 장애물과의 거리를 파악하고 사전에 충돌 회피가 가능하도록 하는 역할을 담당하게 된다. Magnetometer는 3축 지자기 센서로 드론의 방향 정보를 담으며 Lidar 센서 값은 대기 중 물체에 대한 거리를 측정한 값을 나타내어 주변 장애물과의 거리를 파악하고 사전에 충돌 회피가 가능하도록 하는 역할을 담당하게 된다. Command는 전송 데이터의 종류를 정의할 때 사용하며, 메시지를 수신한 드론은 Command 셀의 명령 코드를 확인하여 동작을 인지하게 된다. MSG에는 Command의 명령 코드에 따라 그에 해당하는 전달 메시지가 작성되어 전송되고, CurTime과 Reserve는 각각 현재 시간과 차후 활용을 위해 예약된 셀을 의미한다. 이렇게 GPS, Barometer, Magnetometer, Lidar 등의 센서를 통해 관측된 개별 드론의 상태(state)가 서로 간의 공유될 수 있다면 군집 드론의 공통 목적을 달성하기 위한 협력의 효율이 극대화될 수 있다[21-23].

군집에 속하는 개별 에이전트들은 하나의 주어진 목표를 달성하기 위한 과정에서 부분적인 관측과 제한된 대역폭 안에서 취할 행동을 결정하고 주어진 일을 원활이 해결할 수 있도록 센서를 통해 측정된 정보를 공유하는 것뿐만 아니라 알맞은 통신 프로토콜을 학습해야 한다. 이때 다중 에이전트가 정책을 학습하는 가장 기본적인 구조는 중앙형 학습과 분산형 실행(Centralized Training Decentralized Execution, CTDE) 방식이다. 즉 정책을 학습하는 과정에서는 다른 드론의 상태 및 정보

를 공유하지만 비행하는 과정에서는 해당 정보들이 서로 서로 간에 공유하지 않도록 하는 것이다.



〈그림 13〉 DIAL의 통신 프로토콜 학습 방식[24]

통신 프로토콜을 학습하기 위한 접근 방법인 DIAL (Differentiable Inter-Agent Learning)은 C-Net와 Action Selector, 그리고 DRU 모듈로 구성된다. DIAL 방식에서는 중앙형 학습과 Q-Network의 결합으로 매개변수를 공유할 수 있을 뿐만 아니라 그림 13에 표시된 것과 같이 통신 채널을 통해 한 에이전트에서 다른 에이전트로 그라디언트를 전달할 수 있다는 특징을 갖기 때문에 전체 에이전트들 사이의 end-to-end 학습이 가능해진다. 그라디언트가 한 에이전트로부터 다른 에이전트로 전달되어 흐른다면 더 풍부한 피드백을 서로 간에 제공할 수 있게 되고, 여러 시행착오를 통해 학습하는 것보다 훨씬 적은 학습량을 갖게 되어 더 쉽게 효율적인 통신 프로토콜을 발견할 수 있게 된다. 한 에이전트로부터 다른 에이전트로 전달되는 그라디언트는 보편적인 의미로 학습 과정에서 계속해서 변화하는 값을 나타낸다. 표 2에 정의된 드론의 센서값 역시 비행 과정에서 끊임없이 새로운 값으로 업데이트되는 변화량에 해당하므로, DIAL에서 에이전트 간 전달되는 그라디언트 개념에 표 2의 군집 드론 통신 데이터 포맷을 적용한다면 군집 드론 비행에 최적화된 통신 프로토콜을 학습할 수 있는 효율 및 성능이 향상될 것으로 기대된다.

Ⅳ. 결론

본 논문에서는 효율적인 군집 비행을 실현하기 위한 방법으로 최신 연구되고 있는 드론간의 협력을 통한 다중 에이전트 최신 기술들에 대해 비교하고 분석하였다. 또한, 군집 드론 비행에 적용되기 적합한 강화학습을 기반으로 개별 드론 간 통신 프로토콜을 학습하는 DIAL 다중 에이전트 통신 최적화 기술에 대해 소개하고 군집을 이루는 개별 드론의 상태와 관측한 환경에 대한 정보의 공유를 통해 전체 군집 드론의 능력을 향상시킬 수 있을 것으로 기대하는 관점에서 새로운 연구 방향을 제안하였다.

따라서, 구체적인 데이터 포맷을 기반으로 군집 비행을 수행하는 드론 간에 데이터 교환이 이루어지는 구체적인 대상과 내용을 정의하여, 이를 기반으로 임무 수행 중 개별 드론의 상태와 관측 환경에 대한 정보의 실시간 공유를 통해 보다 빠르고 효율적으로 드론의 군집 비행 임무를 수행할 수 있을 것으로 기대된다. 향후 연구 방향으로는 본 논문에서 제시한 군집 드론 비행의 통신 메시지 포맷에 포함된 구체적인 각종 드론 센서 값들이 DIAL 방식에 적용되어 통신 프로토콜 학습 시에 사용될 수 있도록 DIAL 알고리즘을 수정한 뒤 실험을 진행하고 실험 결과를 토대로 통신 프로토콜의 학습이 향상된 결과를 확인하여 연구의 의미와 결과 분석을 보강할 수 있을 것으로 기대된다.

참고문헌

[1] 구윤표 · 손경환 · 이웅, "스마트 드론을 위한 인공지능 기술 연구 동향," 정보과학회지, 제37권, 제1호, 2019, pp.38-44.

[2] 김중헌, "심층 강화 학습 기술 동향," Broadcasting and Media Magazine, 제27권, 제2호, 2022, pp.26-34.

[3] 문성태 · 최연주 · 김도윤 · 성명훈 · 공현철, "Outdoor Swarm Flight System Based on RTK-GPS," Journal of KIISE, 제43권, 제12호, 2016, pp.1315-1324.

[4] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, M. Riedmiller,"Playing atari with deep reinforcement learning," 2013. https://doi.org/10.48550/arXiv.1312.5602.

[5] 김성현 · 이동훈 · 장인국 · 김현석 · 손영성, "멀티 에이전트 강화학습 연구동향," 정보과학회지, 제37권, 제11호, 2019, pp.8-17.

[6] 변영훈 · 김한솔, "드론 제어를 위한 심층 강화 학습의 최신 연구 사례," 제어로봇시스템학회지, 제28권, 제4호, 2022, pp.12-20.

[7] Jong Ho Bang, So Hyun Byun, Sae Rom Lee, Byung Wook Lee, Tae In Park, Jun Hwang, "A Design of System Architecture and Protocols for Formation Flight of Drone," KSII The 9th International Conference on Internet (ICOIN) Symposium, 2017, pp.315-319.

[8] 문성태 · 김도윤 · 최연주 · 공현철, "실내외 군집 비 행 시스템 소개," 한국지능시스템학회 논문지, 제27권, 제3호, 2017, pp.215-223.

[9] 이덕진, "심층강화학습 기반 환경 인식 및 자율비행," Journal of the KSME, 제59권, 제5호, 2019, pp.43-48.

[10] Joseph Redmon et al, "You Only Look Once: Unified, Real-Time Object Detection," The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp.779-788. https://doi.org./10.1109/CVPR.2016.91.

[11] Bewley, Alex, Zongyuan Ge, Lionel Ott, Fabio Ramos, Ben Upcroft, "Simple online and real time tracking," IEEE International Conference on Image Processing (ICIP), 2016, pp.3464-3468. https://doi.org/10.1109/ICIP.2016.7533003.

[12] Gordon, Daniel, Ali Farhadi, Dieter Fox, "Re 3: Real-Time Recurrent Regression Networks for Visual Tracking of Generic Objects," IEEE Robotics and Automation Letters(RA-L)3, No.2, 2018, pp.788-795.

[13] 유병현 · 데브라니데비 · 김현우 · 송화전 · 박경문 · 이성원, "멀티 에이전트 강화학습 기술 동향," 전자통신동향분석, 제35권, 제6호, 2020, pp.137-149.

[14] J. N. Yasin, M. H. Haghbayan, J. Heikkonen, H. Tenhunen, J. Plosila, "Formation Maintenance and Collision Avoidance in a Swarm of Drones," Proc. of the 2019 3rd International Symposium on Computer Science and Intelligent Control (ISCSIC 2019), Association for Computing Machinery, Article 1, 2019.

[15] Jakob N Foerster, Gregory Farquhar, Triantafy Afouras, "Counterfactual multi-agent policy gradients," Proceedings of the AAAI Conference on Artificial Intelligence, Vol.32, No.1, 2018. https://doi.org/10.1609/aaai.v32i1.11794.

[16] Ryan Lowe, Yi Wu, Aviv Tamar, "Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments," 31st Conference on Neural Information Processing Systems (NIPS 2017), 2017.

[17] Yaodong Yang, Rui Luo, Minne Li, Ming Zhou, "Mean Field Multi-Agent Reinforcement Learning,"ICML, 2018. https://doi.org/10.48550/arXiv.1802.05438.

[18] Shariq Iqbal, Fei Sha, "Actor-Attention-Critic for

Multi-Agent Reinforcement Learning," ICML, 2019.

[19] Hee Chang Ryu, Ha Yong Shin, Jin Kyoo Park, "Multi-Agent Actor-Critic with Hierarchical Graph Attention Network," 34th AAAI Conference on Artificial Intelligence (AAAI-20), 2019.

[20] M. L. Fung, M. Z. Q. Chen, Y. H. Chen, "Sensor fusion: A review of methods and applications," 29th Chinese Control and Decision Conference (CCDC), 2017, pp.3853-3860.

[21] C. V. Goldman, S. Zilberstein, "Decentralized control of cooperative systems: Categorization and complexity analysis," The Journal of Artificial Intelligence Research, Vol.22, 2004, pp.143–174.

[22] E. Pesce, G. Montana, "Improving coordination in small- scale multi-agent deep reinforcement learning through memory-driven communication," Machine Learning, Vol.109, Issue 9-10, 2020, pp.1727-1747. https://doi.org/10.1007/s10994-019-05864-5.

[23] S. Q. Zhang, Q. Zhang, J. Lin, "Efficient communication in multi-agent reinforcement learning via variance based control," in Adv. in Neural Information Processing Systems, 2019, pp.3235-3244.

[24] J. Foerster, I. A. Assael, de N. Freitas, S. Whiteson, "Learning to communicate with deep multi-agent reinforcement learning," In Adv. in Neural Information Processing Systems, 2016, pp.2137-2145.

■ 저자소개 ■

김 은 수
(Kim Eunsu)

2020년 3월~현재
서울여자대학교
소프트웨어융합학과 학부생

관심분야 : 임베디드, 머신러닝, 컴퓨터비전, IoT, 모빌리티, 로보틱스, etc.
E-mail : kimes00@swu.ac.kr

장 연 주
(Jang Yeonju)

2021년 3월~현재
서울여자대학교
소프트웨어융합학과 학부생

관심분야 : 인공지능
E-mail : kssjshyjj@gmail.com

방 정 호
(Bang Jongho)

1990년 중앙대학교 공과대학 전자공학과 (공학사)
1997년 Polytechnic Institute of New York University, Electrical Engineering (공학석사)
2001년 New Jersey Institute of Technology, Electrical Engineering (공학박사)
2002년~2015년
삼성전자 종합기술원 수석연구원
2016년~현재
서울여자대학교 미래산업융합대학 소프트웨어융합학과 부교수

관심분야 : IoT, 무선통신네트워크, 인공지능, 드론, etc.
E-mail : bang6467@swu.ac.kr

논문접수일 : 2024년 7월 16일
수정접수일 : 2024년 8월 06일
게재확정일 : 2024년 8월 16일