(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0051848
(43) 공개일자 2025년04월18일

(51) 국제특허분류(Int. Cl.)
G05D 1/00 (2024.01) *A63F 13/35* (2014.01)
G05D 1/20 (2024.01) *G06N 3/063* (2023.01)
G06N 3/092 (2023.01) *G06T 19/00* (2011.01)
G06T 7/246 (2017.01)
(52) CPC특허분류
G05D 1/0088 (2013.01)
A63F 13/35 (2015.01)
(21) 출원번호 10-2023-0134468
(22) 출원일자 2023년10월10일
심사청구일자 2023년10월10일

(71) 출원인
국립부경대학교 산학협력단
부산광역시 남구 용소로 45, 부경대학교내 (대연동)
(72) 발명자
권기룡
부산광역시 해운대구 좌동순환로 117, 201동 903호 (좌동, 롯데2차아파트)
(74) 대리인
유성원

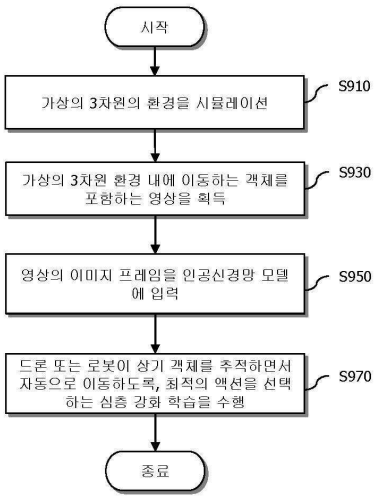
전체 청구항 수 : 총 11 항

(54) 발명의 명칭 **드론 또는 로봇이 객체를 추적하면서 자동으로 이동하도록 인공신경망 모델을 구축하는 방법 및 장치**

(57) 요 약

드론 또는 로봇이 객체를 추적하면서 자동으로 이동하도록 인공신경망 모델을 구축하는 방법이 제시된다. 상기 방법은, 가상의 3차원의 환경을 시뮬레이션하는 단계와; 상기 가상의 3차원 환경 내에 이동하는 객체를 포함하는 영상을 획득하는 단계와; 상기 영상의 이미지 프레임을 인공신경망 모델에 입력하는 단계와; 그리고 상기 드론 또는 로봇이 상기 객체를 추적하면서 자동으로 이동하도록, 상기 드론 또는 로봇의 최적의 액션을 선택하는 심층 강화 학습을 수행함으로써, 상기 인공신경망 모델을 구축하는 단계를 포함할 수 있다.

대 표 도 - 도9



(52) CPC특허분류
G05D 1/0044 (2013.01)
G05D 1/12 (2021.01)
G06N 3/063 (2013.01)
G06N 3/092 (2023.01)
G06T 19/003 (2013.01)
G06T 7/251 (2017.01)
G06T 2207/20084 (2013.01)
G06T 2210/12 (2013.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 1711193874
과제번호 2020-0-01797-004
부처명 과학기술정보통신부
과제관리(전문)기관명 정보통신기획평가원
연구사업명 정보통신방송혁신인재양성
연구과제명 블록체인 기반 및 플랫폼 분야 핵심기술 개발 및 미래 혁신인재 양성
기 여 율 1/2
과제수행기관명 부산대학교산학협력단
연구기간 2023.01.01 ~ 2023.12.31
이 발명을 지원한 국가연구개발사업
과제고유번호 1711193709
과제번호 2016-0-00318-008
부처명 과학기술정보통신부
과제관리(전문)기관명 정보통신기획평가원
연구사업명 정보통신방송혁신인재양성
연구과제명 IoT 및 지능정보 기반 동남권 제조 IT 기술 혁신 및 인재양성
기 여 율 1/2
과제수행기관명 부산대학교산학협력단
연구기간 2023.01.01 ~ 2023.12.31

명 세 서

청구범위

청구항 1

드론 또는 로봇이 객체를 추적하면서 자동으로 이동하도록 인공신경망 모델을 구축하는 방법으로서,

가상의 3차원의 환경을 시뮬레이션하는 단계와;

상기 가상의 3차원 환경 내에 이동하는 객체를 포함하는 영상을 획득하는 단계와;

상기 영상의 이미지 프레임을 인공신경망 모델에 입력하는 단계와;

상기 드론 또는 로봇이 상기 객체를 추적하면서 자동으로 이동하도록, 상기 드론 또는 로봇의 최적의 액션을 선택하는 심층 강화 학습을 수행함으로써, 상기 인공신경망 모델을 구축하는 단계를 포함하는, 방법.

청구항 2

제1항에 있어서, 상기 심층 강화 학습은 상기 드론 또는 로봇의 최적의 액션에 대해서 보상을 주는 것을 포함하는, 방법.

청구항 3

제1항에 있어서, 상기 인공신경망 모델은 상기 객체를 추적하기 위하여, 상기 이미지 프레임 내에서 상기 객체에 대해서 바운딩 박스를 그리는 과정을 포함하는, 방법.

청구항 4

제1항에 있어서, 상기 가상의 3차원의 환경은 게임 엔진 플래폼을 통해서 생성되는, 방법.

청구항 5

제1항에 있어서, 상기 드론 또는 로봇의 최적의 액션은

전진, 후진, 상승, 하강, 좌회전, 우회전, 그리고 정지 중 하나 이상을 포함하는 방법.

청구항 6

드론 또는 로봇이 객체를 추적하면서 자동으로 이동하도록 인공신경망 모델을 구축하기 위한 장치로서,

메모리와;

인공신경망 가속 프로세서를 포함하고, 상기 인공신경망 가속 프로세서는

시뮬레이션되는 가상의 3차원 환경 내에 이동하는 객체를 포함하는 영상의 이미지를 획득하는 제1 회로부와,

상기 드론 또는 로봇이 상기 객체를 추적하면서 자동으로 이동하도록, 상기 드론 또는 로봇의 최적의 액션을 선택하는 심층 강화 학습을 수행함으로써, 상기 인공신경망 모델을 구축하는 제2 회로부를 포함하는 장치.

청구항 7

제1항에 있어서, 상기 심층 강화 학습은 상기 드론 또는 로봇의 최적의 액션에 대해서 보상을 주는 것을 포함하는, 장치.

청구항 8

제1항에 있어서, 상기 인공신경망 모델은 상기 객체를 추적하기 위하여, 상기 이미지 프레임 내에서 상기 객체에 대해서 바운딩 박스를 그리는 과정을 포함하는, 장치.

청구항 9

제1항에 있어서, 상기 가상의 3차원의 환경은 게임 엔진 플래폼을 통해서 생성되는, 장치.

청구항 10

제1항에 있어서, 상기 드론 또는 로봇의 최적의 액션은

전진, 후진, 상승, 하강, 좌회전, 우회전, 그리고 정지 중 하나 이상을 포함하는, 장치

청구항 11

명령어들을 기록하는 컴퓨터 판독가능 저장 매체로서,

상기 명령어들은, 인공신경망 가속 프로세서에 의해 실행될 때, 상기 인공신경망 가속 프로세서로 하여금:

가상의 3차원의 환경을 시뮬레이션하는 과정과;

상기 가상의 3차원 환경 내에 이동하는 객체를 포함하는 영상을 획득하는 과정과;

상기 영상의 이미지 프레임을 인공신경망 모델에 입력하는 과정과;

상기 드론 또는 로봇이 상기 객체를 추적하면서 자동으로 이동하도록, 상기 드론 또는 로봇의 최적의 액션을 선택하는 심층 강화 학습을 수행함으로써, 상기 인공신경망 모델을 구축하는 과정을 수행하도록 하는, 저장 매체.

발명의 설명

기 술 분 야

[0001] 본 개시는 인공신경망 모델을 구축하는 방법 및 장치에 관한 것이다.

배 경 기 술

[0002] 무인항공기는 제조 및 일상 서비스 업무의 모든 분야에서 광범위하게 활용되고 있으며, 드론을 활용하면 공정을 자동화하고 소요 시간을 단축하여 보다 안전한 서비스 활동을 할 수 있다. 예를 들어, 최근 농업 분야에서는 약제 살포 및 시비 과정의 기술 지원에 드론이 최우선적으로 활용되고 있으며, 경작지 맵핑을 통해 경작지의 불규칙한 변화에 대한 중요한 정보를 얻을 수 있다. 대부분의 경우, 드론은 재배자들이 모여 있는 대규모 농장의 생산성을 높이는 데 도움이 될 수 있으며, 보다 효율적인 관수 기술을 가능하게 하는 배수 패턴과 밭의 습-건 지점을 파악하는 데 유용할 수 있다.

[0003] 이 밖에도 드론이 적용되는 분야는 사람이 접근할 수 없는 넓은 지역의 감시 및 제어, AED(Automated Emergency Drones)와 같은 의료용 드론, 수색 및 구조 작업을 위한 자율 드론, UAV 기반 도로 교통 모니터링 애플리케이션, 도시 환경 애플리케이션에서 기상 감지용 드론, 소방관 프로세스를 위한 응급 드론 시스템 애플리케이션, 엔터테인먼트용 드론: 사진, 영화 촬영 등 여러 가지가 있다.

[0004] 그러나, 사용자가 드론을 이용하여 원하는 작업을 수행하기 위해서는, 상당한 훈련 시간과 비용을 요구하는 숙련도가 필요하다.

[0005] 이와 같이 높은 숙련도를 요구하는 부담을 줄이기 위하여, 카메라를 통해 촬영된 영상에서 객체를 추적함으로써, 자동으로 조정할 수 있게 하기 위한 연구들이 진행되고 있다.

[0006] 이러한 연구들은, 드론 뿐만 아니라 로봇, 자율주행차량 등 다방면에서 진행되고 있다.

[0007] 그러나, 실제 물리적 환경 및 조건에서의 정확한 결과를 얻기 위한 객체 추적은 굉장히 어려운 일이고, 많은 시간과 비용이 요구되는 단점이 있다.

발명의 내용

해결하려는 과제

[0008] 따라서, 본 개시는 전술한 문제점을 해결하기 위한 방안들을 제시하는 것을 목적으로 한다.

과제의 해결 수단

[0009] 전술한 목적을 달성하기 위하여, 드론 또는 로봇이 객체를 추적하면서 자동으로 이동하도록 인공신경망 모델을 구축하는 방법이 제시된다. 상기 방법은, 가상의 3차원의 환경을 시뮬레이션하는 단계와; 상기 가상의 3차원 환경 내에 이동하는 객체를 포함하는 영상을 획득하는 단계와; 상기 영상의 이미지 프레임을 인공신경망 모델에 입력하는 단계와; 그리고 상기 드론 또는 로봇이 상기 객체를 추적하면서 자동으로 이동하도록, 상기 드론 또는 로봇의 최적의 액션을 선택하는 심층 강화 학습을 수행함으로써, 상기 인공신경망 모델을 구축하는 단계를 포함할 수 있다.

[0010] 전술한 목적을 달성하기 위하여, 드론 또는 로봇이 객체를 추적하면서 자동으로 이동하도록 인공신경망 모델을 구축하기 위한 장치가 제시된다. 상기 장치는 메모리와; 인공신경망 가속 프로세서를 포함할 수 있다. 상기 인공신경망 가속 프로세서는: 시뮬레이션되는 가상의 3차원 환경 내에 이동하는 객체를 포함하는 영상의 이미지를 획득하는 제1 회로부와, 상기 드론 또는 로봇이 상기 객체를 추적하면서 자동으로 이동하도록, 상기 드론 또는 로봇의 최적의 액션을 선택하는 심층 강화 학습을 수행함으로써, 상기 인공신경망 모델을 구축하는 제2 회로부를 포함할 수 있다.

[0011] 전술한 목적을 달성하기 위하여, 명령어들을 기록하는 컴퓨터 판독가능 저장 매체가 제시된다.

[0012] 상기 명령어들은, 인공신경망 가속 프로세서에 의해 실행될 때, 상기 인공신경망 가속 프로세서로 하여금: 가상의 3차원의 환경을 시뮬레이션하는 과정과; 상기 가상의 3차원 환경 내에 이동하는 객체를 포함하는 영상을 획득하는 과정과; 상기 영상의 이미지 프레임을 인공신경망 모델에 입력하는 과정과; 상기 드론 또는 로봇이 상기 객체를 추적하면서 자동으로 이동하도록, 상기 드론 또는 로봇의 최적의 액션을 선택하는 심층 강화 학습을 수행함으로써, 상기 인공신경망 모델을 구축하는 과정을 수행하도록 할 수 있다.

[0013] 상기 심층 강화 학습은 상기 드론 또는 로봇의 최적의 액션에 대해서 보상을 주는 것을 포함할 수 있다.

[0014] 상기 인공신경망 모델은 상기 객체를 추적하기 위하여, 상기 이미지 프레임 내에서 상기 객체에 대해서 바운딩 박스를 그리는 과정을 포함할 수 있다.

[0015] 상기 가상의 3차원의 환경은 게임 엔진 플래폼을 통해서 생성될 수 있다.

[0016] 상기 드론 또는 로봇의 최적의 액션은 전진, 후진, 상승, 하강, 좌회전, 우회전, 그리고 정지 중 하나 이상을 포함할 수 있다.

발명의 효과

[0017] 본 명세서의 일 개시에 따르면, 객체 추적을 가상 환경에서 수행할 수 있도록 함으로써, 시간과 비용을 절감할 수 있다.

도면의 간단한 설명

[0018] 도 1은 본 명세서에서 제안되는 DRL(Deep Reinforcement Learning) 기반 텐서플로우 에이전트 기반 오브젝트 추적 베이스라인과 게임 엔진의 통합의 일 예를 나타낸 예시도이다.

도 2는 게임 엔진의 DRL 기반 텐서플로우 에이전트 기반 오브젝트 추적 모델의 순서를 나타낸 예시도이다.

도 3은 Q러닝 업데이트 프로세스를 나타낸 예시도이다.

도 4는 양수 A>0 및 음수 A<0 이점을 가진 대리 함수 LCLIP의 확률 비율 r을 나타낸다.

도 5는 DQN 기반 TF-AGENT의 훈련 모델과 PPO 기반 TF-AGENT 훈련 모델에 대한 최소 보상 출력을 나타낸 예시도이고, 도 6은 DQN 기반 TF-AGENT의 훈련 모델과 PPO 기반 TF-AGENT 훈련 모델에 대해 받은 최대 보상 출력을 나타낸 예시도이고, 도 7은 DQN 기반 TF-AGENT의 훈련 모델과 PPO 기반 TF-AGENT 훈련 모델의 평균 보상 수령 결과를 나타낸 예시도이다.

도 8A는 에피소드 50단계에 대한 DQN 기반 TF-AGENT의 테스트 보상 분포를 나타낸 예시도이고, 도 8B는 에피소드 50단계에 대한 PPO 기반 TF-AGENT (b)의 테스트 보상 분포를 나타낸 예시도이다.

도 9는 본 명세서의 일 개시에 따른 일 방법을 나타낸 흐름도이다.

도 10은 본 명세서의 일 개시에 따른 일 장치를 나타낸 예시도이다.

발명을 실시하기 위한 구체적인 내용

[0019] 본 명세서 또는 출원에 개시되어 있는 본 개시의 개념에 따른 실시 예들에 대해서 특정한 구조적 내지 단계적 설명들은 단지 본 개시의 개념에 따른 실시 예를 설명하기 위한 목적으로 예시된 것으로, 본 개시의 개념에 따른 실시 예들은 다양한 형태로 실시될 수 있으며 본 개시의 개념에 따른 실시 예들은 다양한 형태로 실시될 수 있으며 본 명세서 또는 출원에 설명된 실시 예들에 한정되는 것으로 해석되어서는 아니 된다.

[0020] 본 개시의 개념에 따른 실시 예는 다양한 변경을 가할 수 있고 여러 가지 형태를 가질 수 있으므로 특정 실시 예들을 도면에 예시하고 본 명세서 또는 출원에 상세하게 설명하고자 한다. 그러나, 이는 본 개시의 개념에 따른 실시 예를 특정한 개시 형태에 대해 한정하려는 것이 아니며, 본 개시의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.

[0021] 제1 및/또는 제2 등의 용어는 다양한 구성 요소들을 설명하는데 사용될 수 있지만, 상기 구성 요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성 요소를 다른 구성 요소로부터 구별하는 목적으로만, 예컨대 본 개시의 개념에 따른 권리 범위로부터 이탈되지 않은 채, 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소는 제1 구성요소로도 명명될 수 있다.

[0022] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는, 중간에 다른 구성요소가 존재하지 않는 것으로 이해되어야 할 것이다. 구성요소들 간의 관계를 설명하는 다른 표현들, 즉 "~사이에"와 "바로 ~사이에" 또는 "~에 이웃하는"과 "~에 직접 이웃하는" 등도 마찬가지로 해석되어야 한다.

[0023] 본 명세서에서 사용한 용어는 단지 특정한 실시 예를 설명하기 위해 사용된 것으로, 본 개시를 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 서술된 특징, 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0024] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 개시가 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가지고 있다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 의미를 가지는 것으로 해석되어야 하며, 본 명세서에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.

[0025] 실시 예를 설명함에 있어서 본 개시가 속하는 기술 분야에 익히 알려져 있고 본 개시와 직접적으로 관련이 없는 기술 내용에 대해서는 설명을 생략한다. 이는 불필요한 설명을 생략함으로써 본 개시의 요지를 흐리지 않고 더욱 명확히 전달하기 위함이다.

[0026] **I. 객체 추적**

[0027] 최근 UAV 프레임워크와 관련된 많은 혁신과 기술에는 제조 최적화 및 제어 외에도 보안 유지, 제어, 군사 지원, 교통 모니터링 등 여러 가지 이유로 감시 및 표적 추적 장치가 포함된다.

[0028] 현재 혁신적인 기술 개발은 가상 현실 환경 시뮬레이터와 같은 모델링 시스템에서 새로운 아이디어를 얻고 있다. 물리적 객체의 가상 버전과 프로세스 시뮬레이션을 생성하면 적절한 서비스 최적화를 제공하고 어떤 유형의 활동이든 자유롭게 조건을 실험할 수 있다.

[0029] 최근 몇 년 동안 대부분의 UAV 기반 연구 커뮤니티는 CNN(Convolution Neural Network), DNN(Deep Neural Network), LSTM(Long Short Term Memory), RL(Reinforcement Learning) 등과 같은 네트워크 기반 아키텍처와 관련된 여러 신경망을 사용하여 시각 추적 기법의 성능을 향상시키는 데 관심을 기울였다.

[0030] 그러나 이러한 연구들은 공공장소의 여러 카메라에서 촬영한 이미지 컬렉션, 드론 카메라로 촬영한 이미지 및

비디오 세트 등과 같은 물리적 환경 데이터 세트에서 전적으로 결과를 훈련하고 테스트하는 것으로 나타났다.

[0031] 이러한 객체 추적 애플리케이션은 특정 객체 클래스를 사용하여 드론이 정확한 경로에서 동적 객체를 추적하고 대상 추적 프레임워크의 추가 작업으로 특정 메서드를 사용하여 객체를 로컬라이즈하는 방식으로 작동한다.

[0032] 하지만 겉으로 보기에 동일한 측면을 가진 움직이는 물체와 정적인 물체를 추적할 때 어떤 것이 실제 추적 대상인지 인식하는 데는 어려움이 있다.

[0033] 최근의 물체 추적 연구 개발은 가상현실 통합 플랫폼을 통한 집중 학습으로 시작되어 과학자들이 자신의 기술이나 알고리즘을 가상현실 플랫폼과 통합하여 다양한 미세 조정 매개변수 및 조건으로 제안을 테스트할 수 있게 되었다. 이를 통해 연구자들은 하드웨어가 필요 없는 환경에서 비용 부담 없이 자신의 방법을 더 심도 있게 탐구하고 최대한 최적화할 수 있는 더 많은 기회를 얻을 수 있다.

[0034] 일부 에이전트 로봇이나 드론에서 경로를 찾거나 계획하고 화재 환경을 탐색하는 데 사용할 수 있는 심층 Q-러닝 기반 작업이 소방 상황에 대해 제안되었다.

[0035] 이처럼 복잡하고 위험한 상황에서는 다른 구조대원들과 상황 인식을 조율하여 부상자나 피해자를 구조하기 위한 구체적인 계획과 행동으로 상황을 통제하는 것이 시급한 과제이다. 이 프레임워크가 실제 드론이나 로봇에 적용되면 소방관이나 구조대원이 극한의 공포와 혼란스러운 상황에서도 올바른 판단을 내릴 수 있다.

[0036] 본 특허의 발명자는 런타임 환경에서 목표 객체를 추적하기 위해 능동적으로 의사결정을 내릴 수 있는 블록 환경에서 텐서플로우 에이전트 기반 드론 추적 알고리즘을 개발했다. 이 제안에는 시뮬레이션 플랫폼에서 텐서플로우 에이전트 기반 드론을 통한 학습, 추적 및 의사 결정 프로세스를 향상시키기 위한 다양한 보상 기법이 포함되어 있다. 더 높은 정확도를 달성하기 위해 파라미터 값을 올바르게 적용하기 위한 계산적 고려사항이 있으며, 학습 및 테스트 프로세스에서 불필요한 손실과 제약을 제거하기 위해 상태 표현을 공식화했다.

[0037] 본 특허에서는 런타임 환경에서 목표 객체를 추적하기 위해 능동적으로 의사결정을 내릴 수 있는 Blocks 환경에서 tf-agent(TensorFlow-Agent) 기반 드론 추적 알고리즘을 제안한다. 본 특허에서는 시뮬레이션 플랫폼에서 tf-agent 기반 드론을 통한 학습, 추적 및 의사 결정 프로세스를 향상시키기 위한 다양한 보상 기법을 제안한다. 본 특허에서는 더 높은 정확도를 달성하기 위해 파라미터 값을 올바르게 적용하기 위한 계산적 고려사항에 기초하여, 학습 및 테스트 프로세스에서 불필요한 손실과 제약을 제거하기 위해 상태 표현을 공식화하는 것을 제안한다.

[0038] **II. 관련 연구**

[0039] 최근 강화 학습을 통한 객체 추적은 다양한 목표 추적 기법과 통합되어 추적 절차에서 의사 결정을 통해 더 나은 성능을 내면서 발전했다. 하지만 심층 강화 학습(Deep Reinforcement Learning, DRL)을 기반으로 하는 대부분의 시각적 추적 개념은 최종 추정된 목표 좌표가 이상적으로 중앙에 위치하는 동안 검색 영역 내에서 목표 객체를 찾는 방식을 채택한 표현 모델의 경우 성능이 좋지 않은 단점이 있었다.

[0040] **II-A. 심층 강화 학습을 통한 객체 추적**

[0041] 강화 학습을 통한 객체 추적의 발전은 비교적 새로운 아이디어로, 의사 결정 모델과의 객체 로컬라이제이션 및 추적 통합이 학습 및 추적 프로세스에도 적용된다. 여러 연구에 따르면 다양한 환경에서 딥러닝과 RL(Reinforcement Learning)을 결합하면 많은 이점을 얻을 수 있는 것으로 나타났다. 시각적 객체 추적, 시간적 활동 위치 파악, 객체 클래스 식별, 비디오 시퀀스를 통한 객체 인식, 분할(Segmentation) 등은 DRL(Deep Reinforcement Learning)을 사용한 컴퓨터 비전 문제 중 일부에 불과하다. 특히, 최근 몇 년 동안 DRL 프레임워크 연구를 통한 시각적 객체 추적이 증가했는데, 여기서 DRL은 객체 위치를 타겟팅하는 동안 훈련 및 의사 결정 능력을 강화하기 위한 여러 기술과 연관되어 있다. 시각 객체 추적에 DRL을 사용하는 가장 일반적인 방법은 에이전트가 시퀀스의 모든 프레임에서 목표 위치(바운딩 박스)를 추정하고, 정확한 추적 결과를 얻기 위해 최적으로 맞는 동작을 반복적으로 선택해야 한다.

[0042] 따라서 상태 표현은 목표 바운딩 박스가 있는 일반 프레임 상태의 이행 상태이다. 일반적으로 동작은 추적 중 바운딩 박스의 변환 결과이며, 네트워크가 학습 시간 동안 환경에 어떻게 학습하고 적응했는지에 따라 동작이 이동, 스케일링, 전환될 수 있다. DRL 기반 객체 추적에서는 정확도(정밀도)를 보상값으로 볼 수 있으며, 목표 액션 바운딩 박스와 기준값과의 차이를 나타내는데, 일반적으로 이를 IoU(Intersection-over-Union)라고 부른다. 보상 값은 트래킹 정확도를 보여주는 지상 실측 결과와의 액션 값 차이에 따라 변경된다.

[0043] **II-B. 심층 강화 학습을 통한 가상 시뮬레이션 기반 객체 추적**

[0044] 지난 몇 년 동안 대부분의 연구 주제는 모든 유형의 동작, 객체 및 프로세스를 시뮬레이션할 수 있는 가상 시뮬레이션 세계 환경의 혁신적인 트렌드와 상호 작용하여 복잡한 조건의 실험을 통해 결과를 관리하고 최적화할 수 있게 되었다. 시뮬레이션 플랫폼과 알고리즘을 통합하여 동적 및 비동적 동작 모드를 통해 실제 모델에 가까운 시뮬레이션 동작을 활용하면서 테스트 실험을 수행하는 것은 쉬운 일이 아니다.

[0045] 여러 시뮬레이션 플랫폼은 실험 목적으로 작성된 소프트웨어 알고리즘 코드와 통합될 수 있다. 현재 가장 널리 사용되는 오픈 소스 플랫폼은 AirSim 과 Unity로, 가상 세계와 현실 세계 사이의 간극을 메워 자율 제어와 실제 세계의 사실적인 복제본 개발을 지원하고자 한다.

[0046] 두 플랫폼 모두 현재 강화 학습 및 딥 러닝과 같은 데이터 기반 머신 인텔리전스 기술의 개발과 테스트에 긍정적인 영향을 미치기 위해 높은 강도로 기술력을 발전시키고 있다. ConvNet-LSTM function approximator에 드론 에이전트를 적용하여 frame-to-action 전략을 사용하여 목표물의 움직임을 예측하는 심층 강화 학습을 통한 능동 객체 추적 기법이 제안되었다.

[0047] 또한 더 나은 목표 추적 성능을 달성하기 위해 훈련 과정을 강화하기 위한 추가 (ViZDoom 및 언리얼 엔진 시뮬레이션) 환경 증강 기술과 맞춤형 보상 기능을 수행한다.

[0048] 또 다른 가상 시뮬레이션 기반 접근 방식은 감지된 목표 물체를 따라가기 위해 DRL 모델을 통해 단안형 온보드 카메라를 사용하는 방식으로 제안되었다. 이 기법은 미리 계산된 궤적의 여러 센서 정보를 사용하여 가상 환경에서 알고리즘을 채택하는 데 더 정확하고 비용 효율적인 전략을 보여준다고 말한다. 제안한 모델은 학습 과정의 이미지 입력에 대한 경계 상자 정보를 얻기 위해 MobileNet이라는 객체 감지 모델 중 하나와 결합된다. 이 모델에는 네트워크와 알고리즘을 적응적으로 조정하기 위한 융합 기반 탐색 및 활용이 포함된다.

[0049] 또한, 적응적이고 유연한 행동을 하는 RL 에이전트를 제어하기 위해 DQN(Deep Q-Learning) 모델을 사용하여 강화 학습 기반의 드론 행동 객체 추적 프레임워크가 제안되었다. 이 객체 추적 모델에서는 스택 이미지 프레임과 깊이 정보를 입력 프레임으로 적용하여 학습 및 테스트 과정을 진행한다. 제안한 모델은 여러 가지 구조적 변화가 있는 다양한 레벨 환경에서 합리적으로 실험했다. 몇 가지 특정 조건을 사용한 실험 결과는 드론에 대한 RL 기반 기술이 적응 및 일반화 동작에서 성공적임을 보여주었다.

[0050] 최근 연구에서는 심층 강화 학습 알고리즘을 통해 가상 시뮬레이션 기반의 시각 객체 추적 모델을 제안하였는데, 이는 AirSim 드론 에이전트가 런타임 가상 시뮬레이션 환경에서 직접 순차 프레임을 활용하여 목표 객체 클래스를 추적하는 데 채택되었다. 또한, 제안된 모델은 최근 연구 결과물의 성능을 평가하기 위해 공개 데이터 세트로 테스트되었다. 가상 시뮬레이션 플랫폼의 가장 큰 장점은 다양한 미세 조정 기법으로 여러 번 실험을 수행할 수 있고, 높은 정확도로 제안을 개선할 때까지 연구자의 비용이 전혀 들지 않는다는 점이다. 따라서 모델 학습을 강화하고 측위 정확도를 높이기 위해 새로운 가짜 증강 데이터를 생성하거나 공개 세트에서 데이터를 수집/재사용할 필요가 없어 추정 시간과 인적 노력을 줄일 수 있다.

[0051] **III. 본 명세서에서 제안되는 방안**

[0052] 최근의 객체 추적 프레임워크 기술은 제품 배송, 자율 주행 시스템, 보안 시스템(군사 및 운송 포함), 소매 산업, 스마트 시티, 의료 시스템 및 농업 등 제조 및 서비스 산업의 성능 향상에 영향을 미쳤다. 물리적 환경 및 조건에서의 정확한 결과를 얻기 위한 객체 추적은 훨씬 더 어려운 일이다.

[0053] 그러나 시뮬레이션 기술이나 플랫폼을 사용하면 다양한 시뮬레이션 조건과 날씨 변화에서 모델의 성능을 평가하고 확인할 수 있다. 이 연구에서는 Unreal 게임 엔진 시뮬레이션 플랫폼인 Blocks 환경에서 추적 프로세스를 수행하기 위해 강화 학습 기술과 tf-agent(TensorFlow-Agent)를 통합한 대상 추적 방법 중 하나를 제시한다. 이러한 플랫폼의 생산성은 가상 현실 조건에서 가상 드론 에이전트로 실험하고 최상의 성능을 달성하기 위해 미세 조정하는 동안 빛을 볼 수 있다. tf-agent 드론은 간단한 Blocks 환경에서 순차적 프레임을 수신하여 작업, 상태 및 추적을 제어하기 위해 심층 강화 학습 프로세스로 개체 통합을 추적하는 방법을 학습한다. tf-agent는 추적 및 속도의 정확성에 관한 추가 테스트 및 평가를 위해 시뮬레이션 환경에서 주변 환경 및 기존 개체에 적응하기 위해 Blocks 환경에서 훈련된다. 안정성, 속도 및 수치 성능 측면에서 시뮬레이션 프로세스와 통합된 강화 학습 기반 추적기와 같은 일부 기존 방법과 알고리즘을 테스트하고 비교했다.

[0054] 이 기법에서는 학습 및 추적을 포함한 객체 추적 프레임워크의 여러 구성 요소를 포함하는 알고리즘을 만들고 가상 시뮬레이션 환경에서 이를 평가했다. 이 프레임워크의 기본 아이디어는 가상 시뮬레이션 플랫폼의 직접 입

력을 사용하여 동작 공간을 학습하는 것으로, 이를 통해 능동적으로 학습하고 추적하는 데 많은 노력 없이 네트워크를 훈련할 수 있다. 그러나 이 방법은 환경을 학습하고 추적 시퀀스의 각 프레임에서 지속적인 동작 결정을 내리는 데 필요한 객체 특징 정보를 얻기 위해 Q-러닝 네트워크 모델과 올바르게 연결되어야 한다.

[0055] **III-A. 알고리즘 기준선**

[0056] **도 1은 본 명세서에서 제안되는 DRL(Deep Reinforcement Learning) 기반 텐서플로우 에이전트 기반 오브젝트 추적 베이스라인과 게임 엔진의 통합의 일 예를 나타낸 예시도이다.**

[0057] 본 명세서에서 제안하는 방안의 기준선은 도 1과 같다. 먼저, 설계된 알고리즘 모델을 통합하기 위해 필요한 특성 파라미터를 설정하고 AirSim 시뮬레이션 플랫폼을 설치해야 한다. 파이프라인 도 1의 가상 환경 부분에 지정된 가상 시뮬레이션 플랫폼 주위에 정의된 보행 경로를 가진 객체가 항상 시뮬레이션 플랫폼에 삽입되어 있다. 가상 시뮬레이션 플랫폼은 기능 정보와 함께 필수 입력 프레임 시퀀스를 제공한다. 일반, 세그먼트, 그레이스케일(negative) 심도 이미지와 같은 이미지가 동시에 tf-agent DRL 네트워크 레이어를 통해 학습되고 목표 추적 조치를 위한 조치를 취할 수 있다. 이미지 심도를 확보하여 물체의 위치와 목표 클래스를 식별하고 네트워크를 통해 실험하면서 알려지지 않은 조건에 적응할 수 있도록 한다.

[0058] **III-B. 텐서플로우 기반 DRL 객체 추적 모델**

[0059] 가상 시뮬레이션 플랫폼에서 물체를 추적하는 프로세스는 일반적인 최신 타겟 추적 프레임워크 접근 방식과는 상당히 다르다. 목표 물체는 높은 벽, 여러 가지 모양의 물체 등 장애물에 의해 가려진 시뮬레이션 플랫폼 영역에서 자동으로 이동한다. 본 명세서에서는 무작위로 걷는 보행자를 시뮬레이션 환경에 설정하여 가상 에어심 드론 에이전트에 의한 학습 및 추적 조건을 만들었다.

[0060] **도 2는 게임 엔진의 DRL 기반 텐서플로우 에이전트 기반 오브젝트 추적 모델의 순서를 나타낸 예시도이다.**

[0061] 도 2를 참조하여 알 수 있는 바와 같이, 시뮬레이션 플랫폼과 제안한 대안 알고리즘 모델을 통합하여 다양한 조건에서 실험을 통해 대표자를 공동으로 최적화했다. 먼저 환경 시뮬레이션 플랫폼에 일반적인 깊이 이미지와 세그먼테이션 맵을 요청하여 타겟이 있는 픽셀을 얻는다. 다음 단계에서는 가상 시뮬레이션 모델에서 객체를 추가로 인식하기 위해 프레임을 그레이 스케일 및 정규화한다. 타겟이 있는 픽셀을 가져와 생성한 후 바운딩 박스 포인트를 연결하고 다음 프로세스를 위해 네트워크에서 읽을 수 있는 값으로 변환한다.

[0062] 1) DQN-BASED TF-AGENT

[0063] DQN 에이전트는 환경의 확률적 전이에 대한 단순성과 기대치를 위해 결정론적으로 공식화된 이산 행동 공간이 갖는 모든 환경 조건에 적용될 수 있다. 이 모델에서 DQN 에이전트의 주요 목표는 할인된 누적 보상을 최대화하기 위한 정책을 학습하는 것이다:

수학식 1

[0064] $$R_{t_0} = \sum_{t=t_0}^{\infty} \gamma^{t-t_0} r_t$$

[0065] R_{t0}는 반환값이라고도 한다. 대부분의 RL 기반 네트워크에서 할인 계수 ‾는 0과 1 사이에 수렴하는 합계를 강조하는 상수 값이어야 한다. 이를 통해 에이전트는 불확실한 환경 특징 정보를 피함으로써 더 나은 보상 가치 결과를 얻을 수 있고, 상당히 확신할 수 있는 정보보다 관련성이 낮은 정보를 식별할 수 있다. Q^*는 최대 보상을 달성하기 위해 구성된 수학식 2의 결과로 주어진 상태에서 액션을 취했을 때 나오는 리턴을 강조할 수 있는 적절한 보상 또는 리턴 값을 달성하는 것이다.

[0066] $$Q^*: State \times Action \rightarrow \mathbb{R}$$

수학식 2

[0067] $$\pi^*(s) = \underset{a}{\operatorname{argmax}} Q^*(s, a)$$

[0068] 가상 현실 세계 시뮬레이션 모델에서는 Q-network를 생성하고 훈련하기 위해 값 함수와 관련된 상태 및 행동 공간 정보에 액세스할 수 있다. 기본적으로 정책적으로 요구되는 조건의 경우 대부분의 함수는 벨만 방정식으로서, 아래의 수학식 3을 따른다.

수학식 3

[0069] $$Q^{\pi}(s,a) = R + \gamma Q^{\pi}(s',\pi(s'))$$

[0070] 초기 값과 학습된 값의 차이는 등식(Q-값 업데이트 또는 시간차 오차라고도 함)을 따라 계산할 수 있다:

수학식 4

[0071] $$\delta = Q^{new}(s,a) = (1-\alpha)\underbrace{Q(s,a)}_{old\ value} + \alpha\overbrace{\left(R_{t+1} + \gamma \max_{a} Q(s',a')\right)}^{learned\ value}$$

[0072] 위의 수학식 4는 에이전트가 특정 상태-행동 쌍 값을 처음 경험하기 때문에 초기 이전 값은 0이 되는, 이전 값과 학습된 값의 가중치 합과 동일한 것으로 가정하여 시간 단계 t에서 상태-행동쌍의 업데이트 Q값을 계산한다.

[0073] 이전 값에는 (1-α) 가 곱해진다. α학습률은 기본 훈련 네트워크의 경우 α=0.001 로 표시되고 설정됩니다. 새로 계산된 Q값을 덮어쓰는 대신 α 학습률은 초기 상태-행동 쌍에대해 이전에 계산된 Q값의 양을 결정하도록 설정된다. 이후 시간 단계에서 최근에 얻은 Q값을 유지하기 위해 동일한 상태-행동 일치에 대해더 높은 학습률을 부여하여 계산된 Q값에 대해 드론 에이전트를 빠르게 채택한다. 그러나 추가 학습 과정에서 이전 Q값과 새로운 Q값 사이의 균형을 유지하기 위해 균형 잡힌 학습 속도를 유지해야 한다. 학습된 값은 드론 에이전트가 시작 상태 지점에서 무작위로 이동하여 받는 보상 R_{t+1}에 새로운 상태-행동 일치 (s', α') 의 최적 미래 Q값에 대한 할인된 추정치 r를 더한 값으로 간주한다. 학습된 값의 출력에 학습률 α를 곱하여 최적의 정책 값을 업데이트하여 얻는다.

[0074] **도 3은Q러닝 업데이트 프로세스를 나타낸 예시도이다.**

[0075] Q-learning 프로세스 업데이트는 도 3에 나타나 있다.

[0076] 도 3과 같이 에이전트가 가상 환경의 단계를 $Q^{\pi}(s_t, a_t)$을 통해최적의 액션을 선택하고 에이전트의 성과에 대한 보상을 받는 훈련 또는 학습 과정을 통해 여러 액션이 있을 수 있다. 추가 학습 과정을 위해 에이전트는 S_{t+1}상태에서 액션을 선택해 보다 심층적인 특징 결과를 통해 환경을 지속적으로 학습하고 분석한다. 여기서 엡실론 그리드 옵션은 탐색과 탐사 중 무작위로 선택함으로써 탐색과 탐사의 균형을 맞추는 간단한 전략이다. 이 방법에서는 엡실론이 작은 확률로 환경 탐사를 진행할지 아니면 탐사 또는 착취를 선택할 확률이다. 아래 수식에서 엡실론 탐욕적 방법의 수학적 공식으로 행동 선택을 확인할 수 있다:

수학식 5

[0077] $$\text{Action at time (t)} \quad a_t = \begin{cases} \max_{a} Q_t(a) & 1-\epsilon \\ any\ action\ (a) & \epsilon \end{cases}$$

[0078] 추가 학습 과정을 위한 행동 선택 방법은 값이 인 경우 에이전트가 사전 지식을 활용하여 가장 잘 예측된 보상을 활용하고, 그렇지 않은 경우 새로운 최적 옵션을 찾기 위해 탐색을 수행한다.

[0079] 에이전트가 최상의 보상을 생성할 수 있는 행동을 선택하려면 각 행동의 값을 지정해야 한다. 액션 값 추정 함수는 아래의 수학식 6과 같으며, 확률 이론을 사용하여 이러한 값을 정의하는 데 사용된다. 모든 잠재적 행동 목록에서 특정 행동을 선택했을 때 얻을 수 있는 예상 보상을 해당 행동의 가치라고 한다. 따라서 에이전트

(agent)는 특정 액션을 선택했을 때 얻을 수 있는 가치를 알 수 없으므로 '표본 평균' 접근 방식을 사용하여 액션 수행의 가치를 추정한다.

수학식 6

[0080] $$Q_t(a) = \frac{Sum\ of\ rewards\ when\ action(a)taken\ before\ time(t)}{Number\ of\ times\ action(a)taken\ before\ time(t)} = \frac{\sum_{i=1}^{t-1} R_i}{t-1}$$

[0081] 그런 다음 에이전트는 값이 추정되면 추정 값이 가장 큰 작업을 선택하며, 이를 탐욕스러운 작업이라고 한다.

[0082] 2) PPO-BASED TF-AGENT

[0083] 근사 정책 최적화(PPO)는 확률적 기울기 상승을 사용하여 "대리" 목적 함수를 최적화하는 것과 환경과의 상호작용을 통해 데이터를 샘플링하는 것을 번갈아 가며 사용하는 RL 기반 최적화 문제에 대한 간단한 정책 기울기 접근 방식이다. 기본 정책 기울기 방법과 새로운 정책 기울기 방법의 주요 아이디어와 차이점은 표준 모델이 단일 에포크에서 데이터 샘플당 기울기를 업데이트하는 반면, 다중 업데이트 미니 배치 목적 함수가 적용된다는 것이다.

[0084] 강화 학습 에이전트의 정책을 개선하기 위해 근사 정책 최적화(PPO) 알고리즘으로 알려진 정책 그라데이션 기법이 적용된다. PPO는 PPO1과 PPO2를 포함하는 알고리즘 집합이다. 이 제안에서는 PPO1 알고리즘에 초점을 맞출 것이다. 클리핑된 대리 목표는 각 단계에서 정책의 변경량을 제한하여 학습 안정성을 높이기 위한 정책 그라데이션 목표를 대체하는 drop-in 방식이다. 이러한 문제와 기타 어려움을 해결하기 위해 양을 제한할 수 있으며, 정책을 변경하고 항상 개선되도록 지원할 수 있다. 또한 전체 처리 알고리즘과 통합하여 입력에서 효율적인 샘플을 얻고 하이퍼파라미터 튜닝 지표를 최소화하도록 구현할 수 있다. 클리핑된 대리 목표의 기본을 최적화하여 복잡성을 피하면서 동일한 성능 향상을 달성할 수 있다:

수학식 7

[0085] $$L_t^{CLIP}(\theta) = \hat{\mathbb{E}}_t[\min(r_t(\theta)\hat{A}_t, clip(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t)]$$

[0086] $r_t(\theta)\hat{A}_t$ - 최소화 내부에서 이전과 동일한 목표를 식별한다.

[0087] $clip(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t$ - 공식의 이 부분은 동일한 목표이지만 $r(\theta)$ 은 $(1-\epsilon, 1+\epsilon)$ 사이에 잘린다.

[0088] 전체 $\min(r_t(\theta)\hat{A}_t, clip(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t)$ - 에피소드는 이전과 동일한 목적의 최소와 클립된 목적의 최소를 표시한다;

[0089] 대리자 클리핑의 주요 목적은 알고리즘이 너무 욕심을 부려 훈련과 학습 중에 한 번에 너무 많은 것을 업데이트하려고 하는 것을 방지하고, 추정과 요약에 적합한 샘플을 남겨두기 위한 영역 클리핑 프로세스이다. PPO를 사용하면 데이터 스트림에서 해로울 정도로 큰 정책 수정을 트리거하지 않고도 많은 경사도 상승을 수행할 수 있다. 를 통해 수집된 데이터 또는 스트림 데이터를 더 많이 활용하면서 샘플 비효율을 줄일 수 있다. 또한, 실행 중인 PPO 정책은 개별적으로 데이터를 수집하는 N개의 병렬 액터를 사용한다. 데이터는 미니 배치로 수집된 다음 클리핑된 대리 목표 기능을 사용하여 K개의 에포크에 대해 학습된다.

[0090] 클리핑된 대리목표는 에이전트가 수행하는 모든 행동에 영향을 미치고 최적화한다. 경사도 단계가 끝나는 동안 동작이 더 나은(양수) A>0이고 가능성이 더 높으면 업데이트를 중지해야 한다. 그렇지 않으면 드론 동작이 잘못된 지침으로 향하지만 동작이 좋은 경우 리디렉션하거나 초기 상태에서 프로세스를 되돌릴 수 있다. 나쁜 행동(음수) A<0이고 에이전트의 행동으로 얻을 수 있는 결과가 적을 경우, 단계를 단축하거나 행동 공간에서 멀리 이동할 필요가 없으며, 정규화 된 업데이트 수준에 관해서는 균형 잡힌 영역에서 제어하여 최적의 결과를 얻을 수 있다.

$A > 0$ $A < 0$

[0091] **도 4는 양수 및 음수 이점을 가진 대리 함수 L^{CLIP}의 확률 비율 r을 나타낸다.**

[0092] 각 플롯의 빨간색 원은 최적화의 시작점, 즉 r=1을 나타낸다. 대리 함수 L^{CLIP}의 합은 여러 항에 대해 수행된다.

[0093] 위의 도 4는 학습 실행 기간 동안의 대리 목적 함수 최적화 파라메트릭 결과를 확률 비율로 잘라낸 것이다.

[0094] 이 기법은 학습 환경이나 입력 데이터에 따라 정책 변경이 큰 경우에도 모델 에이전트의 확률 최적화를 효과적으로 수행할 수 있다. 클리핑 오브젝트 기법을 적용한 PPO는 유지된 두 개의 정책 네트워크, 즉 $\pi_{\theta k}(a_t|s_t)$ 와 샘플 수집에 마지막으로 사용된 정책 $\pi_{\theta k}(a_t|s_t)$ 사이의 차이를 보여준다. 새로운 정책 평가는 샘플 효율성을 개선하기 위해 이전 정책 샘플을 수집하여 중요한 샘플링에서 비롯된다.

[0095] PPO 액터-크리틱 스타일의 최종 손실 함수는 클리핑된 대리 목표 함수, 값 손실 함수, 엔트로피 보너스를 조합한 방정식은 아래의 수학식 8과 같다:

수학식 8

[0096] $$L_t^{CLIP+VF+S}(\theta) = \hat{\mathbb{E}}_t\left[L_t^{CLIP}(\theta) - c_1 L_t^{VF}(\theta)\right] + c_2 S[\pi_\theta](s_t)$$

[0097] 위의 방정식에는 이해하기 복잡할 수 있는 여러 부분이 포함되어 있지만, 실험 과정에 적용하면서 보다 정확한 결과를 얻는 데 더 우선순위를 두었다. 첫 번째 부분은 수학식 7에서 클리핑된 대리 목적 함수로 설명했다. 수학식 8에서 c_1과 c_2 계산을 위한 관련 파라미터 값의 계수이다.

[0098] $L_t^{VF}(\theta)$ 는 $(V_\theta(s_t) - V_t^{targ})^2$ 의 제곱 오차 값 손실로 식별된다. 가상 환경의 알려지지 않은 어려운 시나리오를 충분히 탐색할 수 있도록 보너스 항목으로 $s[\pi_\theta](s_t)$를 추가했다.

[0099] **IV. 실험 결과 및 고찰**

[0100] 주요 목표와 초점 중 하나는 시뮬레이션 플랫폼의 이점을 활용하여 매개변수 변경이 있는 다양한 조건에서 실험을 수행하는 것이었다. 많은 관련 연구자들이 여러 평가 연구에서 알고리즘을 테스트하고 평가하기 위해 여러 시뮬레이션 플랫폼을 사용했다. 가상 현실 플랫폼의 이점을 얻는 방법에는 여러 가지가 있으며, 대부분의 경우 실험 목적으로만 사용되지만 학습, 훈련 및 테스트에서도 더 광범위하게 우선순위를 지정할 수 있다. 유니티, 언리얼 엔진, 세시움과 같은 시뮬레이션 플랫폼의 최신 개발은 여러 가지 비현실적인 상황에서 최첨단 모델을 처리하고 실험할 수 있는 큰 기회와 이점을 제공한다. 시뮬레이터의 주요 기능 중 하나는 프로그래밍 언어(Javascript, Python, Go, Java, Kotlin, PHP, C#, Swift 등)와 프레임워크(Angular, jQuery, React, Ruby 및 Rails, Vue, ASP.NET Core, Django, Express 등) 간의 상호연결이다. 그러나 이러한 아키텍처와 프레임워크를 구축하거나 설정하는 것은 매우 어렵고, 타사 프로그램과 라이브러리의 충돌 및 불균형으로 인해 모든 경우에 성공할 수 없다. 이 연구 작업에서는 다양한 파라미터 변경 및 미세 조정을 통해 여러 실험을 수행했으며, 다음 절에서 설명하기로 한다.

[0101] **IV-A.트레이닝 결과**

[0102] 환경 공간을 학습하고 향후 테스트 및 평가 목적으로 모델을 생성하기 위해 무작위로 움직이는 물체를 삽입하여 간단한 블록 환경으로 제안 모델을 학습시켰다. 0.001의 학습률 구성으로 보다 비교 가능한 출력 결과를 얻기 위해 DQN 기반 TF-AGENT와 PPO 기반 TF-AGENT 두 가지 유형의 모델을 적용했다.

[0103] **도 5는 DQN 기반 TF-AGENT의 훈련 모델과 PPO 기반 TF-AGENT 훈련 모델에 대한 최소 보상 출력을 나타낸 예시도이고, 도 6은 DQN 기반 TF-AGENT의 훈련 모델과 PPO 기반 TF-AGENT 훈련 모델에 대해 받은 최대 보상 출력을 나타낸 예시도이고, 도 7은 DQN 기반 TF-AGENT의 훈련 모델과 PPO 기반 TF-AGENT 훈련 모델의 평균 보상 수령 결과를 나타낸 예시도이다.**

[0104] 도 5를 참조하면, 일반 블록 환경에서 학습된 모델들의 최소 달성 보상 산출값이 나타나 있다. DQN 기반 TF-AGENT는 파란색으로, PPO 기반 TF-AGENT 모델은 분홍색으로 표시되어 있다. 최소 보상은 에이전트가 학습 과정

에서 보상으로 받을 수 있는 최소값이다. 대부분의 문제에는 차선책에 대한 페널티가 포함되므로 최소 보상은 일반적으로 음수 값으로 설정된다. 훈련 기간은 2000으로, 보상은 50으로 각각 설정했다. 그리고 각 방법의 배경은 학습 에이전트의 전반적인 성능을 보여주기 위해 플롯 탭 색상으로 설정되었다. 각 에이전트 모델에 따라 초기에는 DQN 기반 에이전트가 더 나은 성능을 보이지만, 훈련 에포크가 끝날 무렵에는 PPO 기반 에이전트가 DQN 기반 모델 에이전트보다 더 나은 결과를 얻는 것을 알 수 있다.

[0105] 도 6을 참조하면, 훈련 에포크와 보상을 각각 2000과 50으로 설정했을 때 받은 최대 훈련 보상 성과가 나와 있다. 최대 보상은 agent가 훈련 과정에서 받을 수 있는 가장 높은 값이다. 대부분의 문제에는 최적의 의사 결정에 대한 보상이 포함되므로 최대 보상은 일반적으로 양수 값이다. 이 그래프에서는 DQN 기반 TF-AGENT 모델이 초기에는 더 높은 보상 값을 얻지만, 훈련 단계가 끝날 때까지 400개의 에포크가 지나면 PPO 기반 모델의 성능이 더 좋아진다. 가능한 보상 범위를 이해하면 학습률이나 할인 계수와 같은 모델의 하이퍼파라미터를 설정하는 데 유용할 수 있다. 또한 에이전트가 획득한 보상을 가능한 최소값과 최대값과 비교할 수 있으므로 훈련된 에이전트의 성능을 평가하는 데 도움이 될 수 있다.

[0106] 도 7을 참조하면, 평균 보상은 agent가 환경과 상호작용하는 동안 DQN 및 PPO 기반 모델 알고리즘을 사용하여 학습하는 동안 받은 보상의 평균값이 나타나 있다. 평균 보상은 훈련 중 에이전트의 성능을 평가하는 중요한 지표이다. 에이전트는 학습 과정에서 시간이 지남에 따라 얻은 누적 보상을 최대화하는 최적의 정책을 학습하려고 노력한다. 평균 평가 보상은 모든 에피소드 동안 받은 보상의 합계를 총 에피소드 수로 나누어 계산할 수 있다. 이를 통해 agent는 학습된 정책을 사용하여 환경과 상호작용할 때 평균적으로 받게 될 예상 보상을 추정할 수 있다.

[0107] 받은 평균 보상을 평가하면 다양한 구성에서 DQN과 PPO 기반 모델의 성능 차이를 확인할 수 있다. 하지만 일부 시나리오에서는 평균 보상이 agent의 성과를 평가하기 위한 가장 적합한 지표가 아닐 수도 있다.

[0108] **IV-B.테스트 결과**

[0109] 제안한 DQN과 PPO 기반 모델 에이전트를 동일한 환경 조건이지만 보이지 않는 다른 테스트 에피소드로 테스트하여 모델의 능력을 탐색하고 성능을 비교했다. 앞서 언급했듯이 DRL 기반 알고리즘의 성능 평가는 성능 지표 평가 및 비교 기법에서 다른 최신 알고리즘과 차이가 있다. 에이전트 기반 모델의 정확도는 받은 보상 값으로 보거나 확인할 수 있다. 앞서 언급했듯이 에이전트가 학습 중에 얻은 평균 보상을 사용하여 모델을 생성하고 이 모델을 성능 지표로 테스트 프로세스에 적용할 수 있다. 이 메트릭은 에이전트가 환경을 탐색하고 예상되는 결과 추적을 얻는 능력을 측정한다. 아래 다이어그램(도 8)은 보이지 않는 테스트 시나리오에서 받은 보상의 백분율과 함께 DQN 기반 TF-AGENT 모델의 테스트 결과를 나타낸다. 테스트 세션에서는 50단계의 에피소드에서 받은 보상 비율을 100으로 설정한다.

[0110] **도 8A는 에피소드 50단계에 대한 DQN 기반 TF-AGENT의 테스트 보상 분포를 나타낸 예시도이고, 도 8B는 에피소드 50단계에 대한 PPO 기반 TF-AGENT (b)의 테스트 보상 분포를 나타낸 예시도이다.**

[0111] 열과 빨간색 선으로 표시된 각 단계에서 받은 전체 보상은 도 8A에서 일반 보상으로 학습 및 테스트한 DQN 기반 TF-AGENT 테스트 결과의 평활화 값을 나타낸다.

[0112] 도 8B는 50단계의 에피소드에서 PPO 기반 TF-AGENT가 받은 보상 비율 테스트 결과를 빨간색 선으로 매끄럽게 처리한 결과와 함께 보여준다. 모든 훈련 단계에서 보상을 받은 DQN 모델과 PPO 모델 간의 결과가 다르다. 테스트 결과에서 볼 수 있듯이 두 모델 모두 모든 테스트 단계에서 정확도가 높고 명확한 학습을 통해 높은 결론을 도출한 경우 높은 성능을 보였다.

[0113] **V. 결론**

[0114] 본 특허에서는 제안한 알고리즘의 성능을 실험하고 평가하기 위해 간단한 blocks 환경과 통합한 DQN 및 PPO 기반의 TF-AGENT 모델 기반 객체 추적 프레임워크를 제시했다. 알고리즘의 전반적인 성능을 보여주기 위해 시뮬레이션 플랫폼과 통합하였다. 시뮬레이션 플랫폼은 전반적인 상태를 실험하고 평가하기 위해 필수적인 세 가지 유형의 입력 이미지를 제공한다. 가상의 드론 에이전트를 통해 가상의 현실적인 조건에서 실험하고, 최상의 성능 또는 원하는 성능을 달성하기 위한 미세 조정을 수행하면서 플랫폼의 생산성을 확인할 수 있다. 이 제안에서 tf-agent 드론은 간단한 blocks 환경에서 순차적으로 프레임을 수신하여 동작, 상태, 추적 등을 제어하는 심층 강화학습 프로세스를 통해 객체 통합 추적 방법을 학습한다. TF-agent는 blocks 환경에서 주변 환경과 기존 사물에 대한 적응을 위해 학습하고, 시뮬레이션 환경에서 추적 정확도와 속도에 대한 추가 테스트 및 평가를 진행한다. 시뮬레이션 프로세스와 통합된 강화학습 기반 트래커와 같은 기존 방법과 안정성, 속도, 수치적 성능 측

면에서 알고리즘을 테스트하고 비교했다.

[0115] **VI. 본 명세서의 개시들의 정리**

[0116] **도 9는 본 명세서의 일 개시에 따른 일 방법을 나타낸 흐름도이다.**

[0117] 도 9를 참조하면, 가상의 3차원의 환경이 시뮬레이션될 수 있다(S910).

[0118] 상기 가상의 3차원의 환경은 게임 엔진 플래폼을 통해서 시뮬레이션될 수 있다.

[0119] 다음으로, 가상의 3차원 환경 내에 이동하는 객체를 포함하는 영상이 획득될 수 있다(S930).

[0120] 그러면, 영상의 이미지 프레임이 인공신경망 모델에 입력될 수 있다(S950).

[0121] 드론 또는 로봇이 상기 객체를 추적하면서 자동으로 이동하도록, 최적의 액션을 선택하는 심층 강화 학습을 수행됨으로써, 인공신경망 모델이 구축될 수 있다(S970). 상기 드론 또는 로봇의 최적의 액션은 전진, 후진, 상승, 하강, 좌회전, 우회전, 그리고 정지 중 하나 이상을 포함할 수 있다.

[0122] 상기 심층 강화 학습은 상기 드론 또는 로봇의 최적의 액션에 대해서 보상을 주는 것을 포함할 수 있다.

[0123] 상기 인공신경망 모델은 상기 객체를 추적하기 위하여, 상기 이미지 프레임 내에서 상기 객체에 대해서 바운딩 박스를 그리는 과정을 포함할 수 있다.

[0124] **도 10은 본 명세서의 일 개시에 따른 일 장치를 나타낸 예시도이다.**

[0125] 도 10을 참조하여 알 수 있는 바와 같이, 포인트 클라우드를 처리하기 위한 장치(100)가 제시된다. 상기 장치(100)는 인공신경망 가속 프로세서(110)와 메모리(120)와 일반 프로세서(130)를 포함할 수 있다.

[0126] 한편, 상기 인공신경망 가속 프로세서(110)는 GPU(Graphic Processing Unit)이거나,NPU(Neural Processing Unit), 그래프 처리 유닛(graph processing unit), 혹은 TPU(Tensor Processing Unit)으로 불리기도 한다.

[0127] 상기 인공신경망 가속 프로세서(110)는 복수의 회로부를 포함할 수 있다.

[0128] 예를 들어, 도시된 바와 같이 상기 인공신경망 가속 프로세서(110)는 시뮬레이션되는 가상의 3차원 환경 내에 이동하는 객체를 포함하는 영상의 이미지를 획득하는 제1 회로부를 포함할 수 있다.

[0129] 또한 예를 들어, 도시된 바와 같이 상기 인공신경망 가속 프로세서(110)는 상기 드론 또는 로봇이 상기 객체를 추적하면서 자동으로 이동하도록, 상기 드론 또는 로봇의 최적의 액션을 선택하는 심층 강화 학습을 수행함으로써, 상기 인공신경망 모델을 구축하는 제2 회로부를 포함할 수 있다.

[0130] 상기 메모리(120)는 휘발성 메모리 또는 비휘발성 메모리일 수 있다. 상기 휘발성 메모리는 DRAM일 수 있다. 상기 비휘발성 메모리는 플래쉬 메모리일 수 있다.

[0131] 상기 메모리(120)는 상기 인공신경망 가속 프로세서(110)를 동작시키기 위한 명령어들을 저장할 수 있다. 상기 명령어들은, 인공신경망 가속 프로세서(110)에 의해실행될 때, 상기 인공신경망 가속 프로세서(110)로 하여금: 시뮬레이션되는 가상의 3차원 환경 내에 이동하는 객체를 포함하는 영상의 이미지를 획득하는 과정과; 그리고 상기 드론 또는 로봇이 상기 객체를 추적하면서 자동으로 이동하도록, 상기 드론 또는 로봇의 최적의 액션을 선택하는 심층 강화 학습을 수행함으로써, 상기 인공신경망 모델을 구축하는 과정을 수행하도록 할 수 있다.

[0132] 이와 같이, 상기 메모리(120)에 저장되는 상기 명령어들은 상기 인공신경망 가속 프로세서(110)에 포함된 복수의 회로부들이 각 동작 별로 전술한 제1 회로부, 제2 회로부, 제3 회로부 등으로 구분되도록 할 수 있다. 상기 메모리(120)에 저장되는 상기 명령어들이 실행되지 않는 다면, 상기 인공신경망 가속 프로세서(110)에 포함된 복수의 회로부들은 전술한 제1 내지 제3 회로부 등으로 기능적으로 구분되지 못할 수도 있다. 다시 말해서, 상기 메모리(120)에 저장되는 상기 명령어들이 상기 인공신경망 가속 프로세서(110)에 의해서 실행되면, 상기 복수의 회로부들이 전술한 제1 내지 제3 회로부 등으로 기능적으로 구분될 수 있게 되는 것이다.

[0133] 상기 일반 프로세서(130)는 RISC(reduced instruction set computer) 기반의 아키텍처로 만들어진 프로세서 또는 CISC(complex instruction set computer) 기반의 아키텍처로 만들어진 프로세서일 수 있다. 예를 들어, 상기 일반 프로세서는 Intel™의 CPU(Central processing unit)이거나 AMD™의 CPU 혹은 ARM™의 CPU일 수 있다.

[0134] 그 밖에 자세한 내용은 전술한 내용과 동일하므로,중복하여 설명하지 않기로 한다.

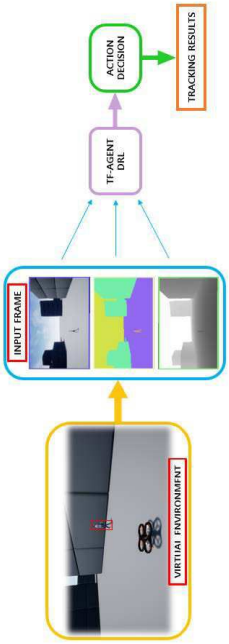
[0135] 본 명세서와 도면에 나타난 본 개시의 예시들은 본 개시의 기술 내용을 쉽게 설명하고 본 개시의 이해를 돕기

위해 특정 예를 제시한 것뿐이며, 본 명의 범위를 한정하고자 하는 것은 아니다. 지금까지 설명한 예시들 이외에도 다른 변형 예들이 실시 가능하다는 것은 본 개시가 속하는 기술 분야에서 통상의 지식을 가진 자에게 자명한 것이다.

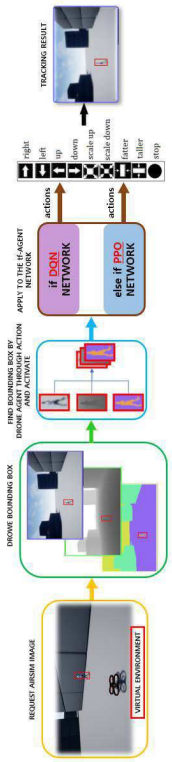
[0136] 본 명세서에 기재된 청구항들은 다양한 방식으로 조합될 수 있다. 예를 들어, 본 명세서의 방법 청구항의 기술적 특징이 조합되어 장치로 구현될 수 있고, 본 명세서의 장치 청구항의 기술적 특징이 조합되어 방법으로 구현될 수 있다. 또한, 본 명세서의 방법 청구항의 기술적 특징과 장치 청구항의 기술적 특징이 조합되어 장치로 구현될 수 있고, 본 명세서의 방법 청구항의 기술적 특징과 장치 청구항의 기술적 특징이 조합되어 방법으로 구현될 수 있다.

도면

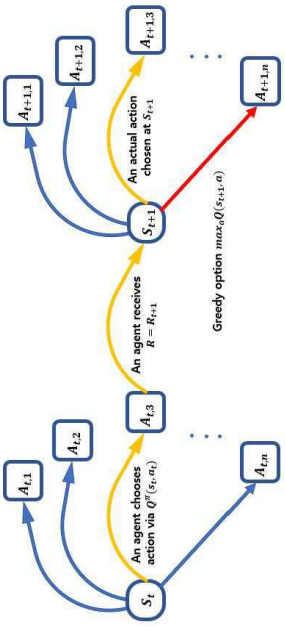
도면1



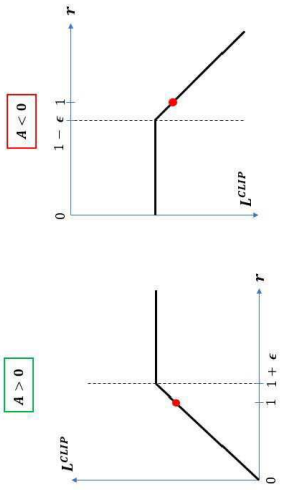
도면2


TRACKING RESULT
right
left
up
down
scale up
scale down
fatter
taller
stop
actions
actions
APPLY TO THE R-AGENT NETWORK
if DQN NETWORK
else if PPO NETWORK
FIND BOUNDING BOX BY DRONE AGENT THROUGH ACTION PREDICTION
DRONE BOUNDING BOX
REQUEST AIRSIM IMAGE
VIRTUAL ENVIRONMENT

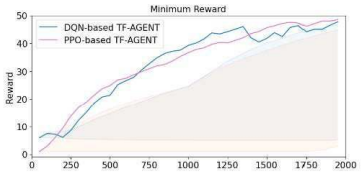
도면3


$A_{t+1,1}$
$A_{t+1,2}$
$A_{t+1,3}$
$A_{t+1,n}$
An actual action chosen at S_{t+1}
S_{t+1}
An agent receives $R = R_{t+1}$
Greedy option $max_a Q(S_{t+1}, a)$
$A_{t,1}$
$A_{t,2}$
$A_{t,3}$
$A_{t,n}$
An agent chooses action via $Q^{\pi}(s_t, a_t)$
S_t

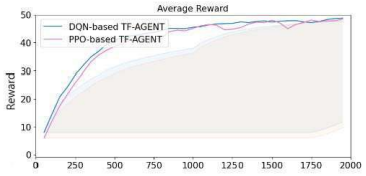
도면4



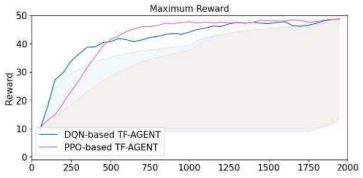
도면5



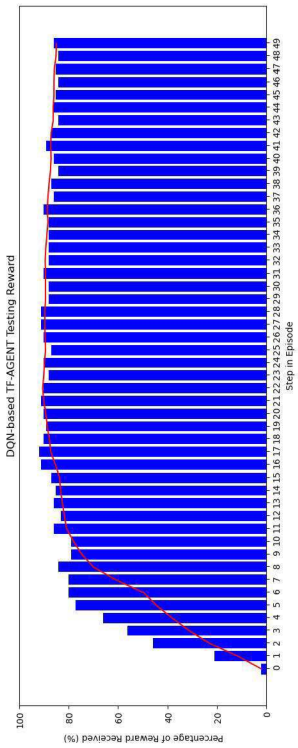
도면6



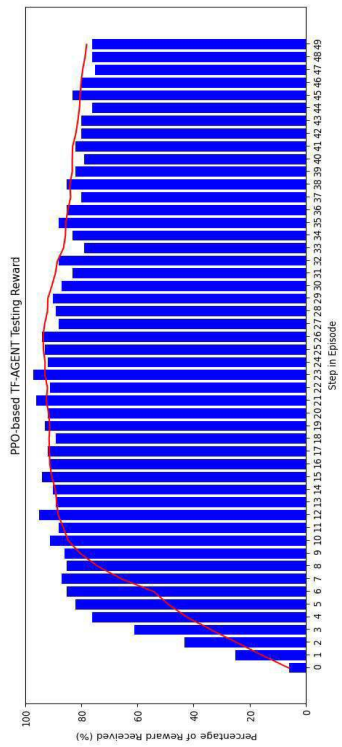
도면7



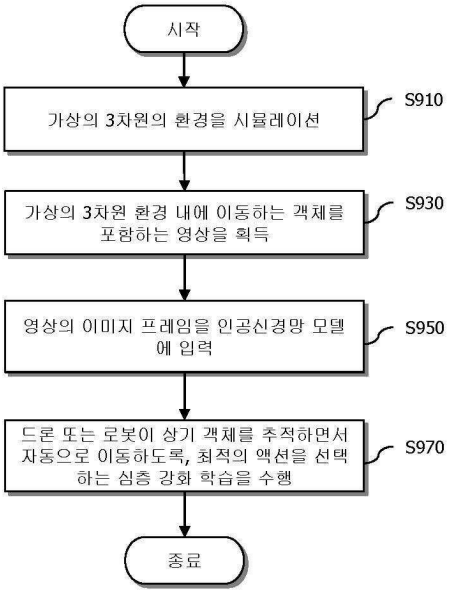
도면8a


DQN-based TF-AGENT Testing Reward
Step in Episode
Percentage of Reward Received (%)

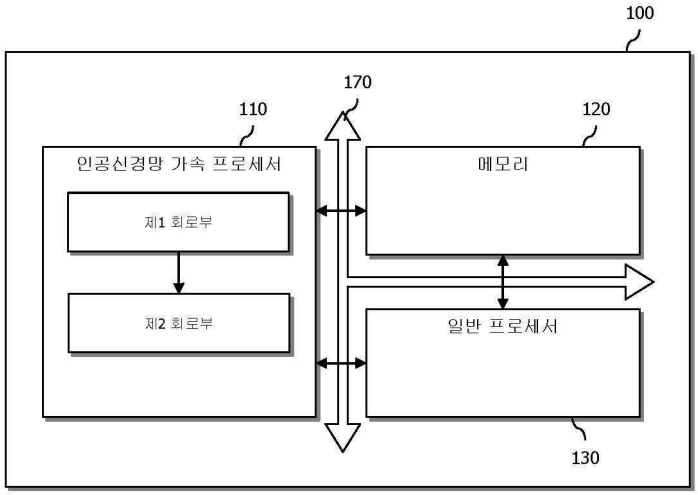
도면8b


PPO-based TF-AGENT Testing Reward
Step in Episode
Percentage of Reward Received (%)

도면9


시작
가상의 3차원의 환경을 시뮬레이션
S910
가상의 3차원 환경 내에 이동하는 객체를 포함하는 영상을 획득
S930
영상의 이미지 프레임을 인공신경망 모델에 입력
S950
드론 또는 로봇이 상기 객체를 추적하면서 자동으로 이동하도록, 최적의 액션을 선택하는 심층 강화 학습을 수행
S970
종료

도면10


100
170
110
120
인공신경망 가속 프로세서
제1 회로부
제2 회로부
메모리
일반 프로세서
130