(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2024년06월19일
(11) 등록번호 10-2675973
(24) 등록일자 2024년06월12일

(51) 국제특허분류(Int. Cl.)
G10L 15/16 (2006.01) *G06T 19/00* (2011.01)
G06T 7/50 (2017.01) *G10L 15/183* (2013.01)
G10L 15/22 (2006.01)
(52) CPC특허분류
G10L 15/16 (2013.01)
G06T 19/00 (2013.01)
(21) 출원번호 10-2023-0183802
(22) 출원일자 2023년12월15일
심사청구일자 2023년12월15일
(56) 선행기술조사문헌
KR1020230085744 A
(뒷면에 계속)

(73) 특허권자
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
광주과학기술원
광주광역시 북구 첨단과기로 123 (오룡동)
(72) 발명자
최종현
서울특별시 서대문구 증가로 191, 101동 1603호 (남가좌동, DMC래미안클라시스)
김병휘
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
(뒷면에 계속)
(74) 대리인
특허법인 수

전체 청구항 수 : 총 20 항 심사관 : 경연정

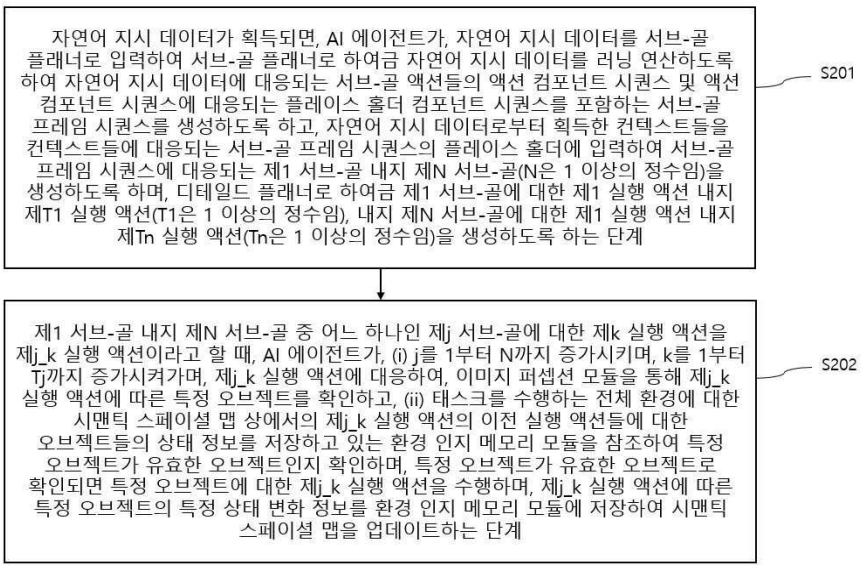
(54) 발명의 명칭 **컨텍스트 인지 플래닝 모듈 및 환경 인지 메모리 모듈을 포함하는 CAPEAM 모델에 따른 태스크 수행 방법 및 이를 사용한 AI 에이전트**

(57) 요 약

본 발명은 컨텍스트 인지 플래닝 모듈 및 환경 인지 메모리 모듈을 포함하는 CAPEAM 모델에 따른 태스크 수행 방법에 관한 것으로서, 보다 상세하게는 (a) 자연어 지시 데이터가 획득되면, AI 에이전트가, 상기 자연어 지시 데이터를 서브-골 플래너로 입력하여 상기 서브-골 플래너로 하여금 상기 자연어 지시 데이터를 러닝 연산하도록

(뒷면에 계속)

대 표 도 - 도2



하여 상기 자연어 지시 데이터에 대응되는 서브-골 액션들의 액션 컴포넌트 시퀀스 및 상기 액션 컴포넌트 시퀀스에 대응되는 플레이스 홀더 컴포넌트 시퀀스를 포함하는 서브-골 프레임 시퀀스를 생성하도록 하고, 상기 자연어 지시 데이터로부터 획득한 컨텍스트들을 상기 컨텍스트들에 대응되는 상기 서브-골 프레임 시퀀스의 플레이스 홀더에 입력하여 상기 서브-골 프레임 시퀀스에 대응되는 제1 서브-골 내지 제N 서브-골 - 상기 N은 1 이상의 정수임- 을 생성하도록 하며, 디테일드 플래너로 하여금 상기 제1 서브-골에 대한 제1 실행 액션 내지 제T1 실행 액션 - 상기 T1은 1 이상의 정수임 -, 내지 상기 제N 서브-골에 대한 제1 실행 액션 내지 제Tn 실행 액션 - 상기 Tn은 1 이상의 정수임 - 을 생성하도록 하는 단계; 및 (b) 상기 제1 서브-골 내지 상기 제N 서브-골 중 어느 하나인 제j 서브-골에 대한 제k 실행 액션을 제j_k 실행 액션이라고 할 때, 상기 AI 에이전트가, (i) 상기 j를 1부터 상기 N까지 증가시키며, 상기 k를 1부터 상기 Tj까지 증가시켜가며, 상기 제j_k 실행 액션에 대응하여, 이미지 퍼셉션 모듈을 통해 상기 제j_k 실행 액션에 따른 특정 오브젝트를 확인하고, (ii) 상기 태스크를 수행하는 전체 환경에 대한 시맨틱 스페이셜 맵 상에서의 상기 제j_k 실행 액션의 이전 실행 액션들에 대한 오브젝트들의 상태 정보를 저장하고 있는 상기 환경 인지 메모리 모듈을 참조하여 상기 특정 오브젝트가 유효한 오브젝트인지 확인하며, 상기 특정 오브젝트가 유효한 오브젝트로 확인되면 상기 특정 오브젝트에 대한 상기 제j_k 실행 액션을 수행하며, 상기 제j_k 실행 액션에 따른 상기 특정 오브젝트의 특정 상태 변화 정보를 상기 환경 인지 메모리 모듈에 저장하여 상기 시맨틱 스페이셜 맵을 업데이트하는 단계를 포함하는 방법이 개시된다.

(52) CPC특허분류
G06T 7/50 (2017.01)
G10L 15/183 (2013.01)
G10L 15/22 (2013.01)
G10L 2015/228 (2013.01)

(72) 발명자
김진연
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
민철홍
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
김유영
제주특별자치도 제주시 연화로 25, 202동 1301호 (연동, 대림2차아파트)

(56) 선행기술조사문헌
US20120110579 A1
Suvvansh Bhambri, Byeonghwi Kim, and Jonghyun Choi.Multi-level compositional reasoning for interactive instruction following. In AAAI, 2023.
Wenlong Huang et al., Language models as zero-shot planners: Extracting actionable knowledge for embodied agents. In ICML, 2022
Shreyas Sundara Raman et al., Planning with large language models via corrective re-prompting. In Foundation Models for Decision Making Workshop NeurIPS, 2022
US20210380128 A1

이 발명을 지원한 국가연구개발사업

과제고유번호	1711193986
과제번호	2020-0-01361-004
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신방송혁신인재양성
연구과제명	인공지능대학원지원(연세대학교)
기 여 율	20.00/100
과제수행기관명	연세대학교 산학협력단
연구기간	2023.01.01 ~ 2023.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711193817
과제번호	2022-0-00113-002
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	사람중심인공지능핵심원천기술개발
연구과제명	지속 가능한 협업형 멀티 모달 평생 학습 프레임워크 개발
기 여 율	40.00/100
과제수행기관명	연세대학교 산학협력단
연구기간	2023.01.01 ~ 2023.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711193448
과제번호	2022-0-00871-002
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	사람중심인공지능핵심원천기술개발
연구과제명	(1세부) 인공지능 에이전트 협업기반 신경망 변이 및 지능 강화 기술 개발
기 여 율	40.00/100
과제수행기관명	한국전자통신연구원
연구기간	2023.01.01 ~ 2023.12.31

공지예외적용 : 있음

명 세 서

청구범위

청구항 1

컨텍스트 인지 플래닝 모듈 및 환경 인지 메모리 모듈을 포함하는 CAPEAM 모델에 따른 태스크 수행 방법에 있어서,

(a) 자연어 지시 데이터가 획득되면, AI 에이전트가, 상기 자연어 지시 데이터를 서브-골 플래너로 입력하여 상기 서브-골 플래너로 하여금 상기 자연어 지시 데이터를 러닝 연산하도록 하여 상기 자연어 지시 데이터에 대응되는 서브-골 액션들의 액션 컴포넌트 시퀀스 및 상기 액션 컴포넌트 시퀀스에 대응되는 플레이스 홀더 컴포넌트 시퀀스를 포함하는 서브-골 프레임 시퀀스를 생성하도록 하고, 상기 자연어 지시 데이터로부터 획득한 컨텍스트들을 상기 컨텍스트들에 대응되는 상기 서브-골 프레임 시퀀스의 플레이스 홀더에 입력하여 상기 서브-골 프레임 시퀀스에 대응되는 제1 서브-골 내지 제N 서브-골 - 상기 N은 1 이상의 정수임- 을 생성하도록 하며, 디테일드 플래너로 하여금 상기 제1 서브-골에 대한 제1 실행 액션 내지 제T1 실행 액션 - 상기 T1은 1 이상의 정수임 -, 내지 상기 제N 서브-골에 대한 제1 실행 액션 내지 제Tn 실행 액션 - 상기 Tn은 1 이상의 정수임 - 을 생성하도록 하는 단계; 및

(b) 상기 제1 서브-골 내지 상기 제N 서브-골 중 어느 하나인 제j 서브-골에 대한 제k 실행 액션을 제j_k 실행 액션이라고 할 때, 상기 AI 에이전트가, (i) 상기 j를 1부터 상기 N까지 증가시키며, 상기 k를 1부터 상기 Tj까지 증가시켜가며, 상기 제j_k 실행 액션에 대응하여, 이미지 퍼셉션 모듈을 통해 상기 제j_k 실행 액션에 따른 특정 오브젝트를 확인하고, (ii) 상기 태스크를 수행하는 전체 환경에 대한 시맨틱 스페이셜 맵 상에서의 상기 제j_k 실행 액션의 이전 실행 액션들에 대한 오브젝트들의 상태 정보를 저장하고 있는 상기 환경 인지 메모리 모듈을 참조하여 상기 특정 오브젝트가 유효한 오브젝트인지 확인하며, 상기 특정 오브젝트가 유효한 오브젝트로 확인되면 상기 특정 오브젝트에 대한 상기 제j_k 실행 액션을 수행하며, 상기 제j_k 실행 액션에 따른 상기 특정 오브젝트의 특정 상태 변화 정보를 상기 환경 인지 메모리 모듈에 저장하여 상기 시맨틱 스페이셜 맵을 업데이트하는 단계;

를 포함하는 방법.

청구항 2

제1항에 있어서,

상기 (b) 단계에서,

상기 AI 에이전트가, 상기 이미지 퍼셉션 모듈로부터 획득된 상기 전체 환경에 대한 공간 정보를 참조로 하여 상기 전체 환경에 대응하는 뎁스-맵(depth map)을 생성하고, 상기 이미지 퍼셉션 모듈로부터 상기 전체 환경 내의 전체 오브젝트 중 적어도 일부 각각에 대응하는 전체 오브젝트 마스크 각각을 획득하며, 상기 전체 오브젝트 마스크 각각과 상기 뎁스-맵을 3D 세계 좌표로 백프로젝팅(back-projecting)하여 상기 시맨틱 스페이셜 맵을 구축하고, 상기 환경 인지 메모리 모듈에 저장된 상기 특정 상태 변화 정보 - 상기 특정 상태 변화 정보는, 상기 제j_k 실행 액션을 수행하기 이전에 획득된 상기 특정 오브젝트의 특정 최초 위치 정보, 상기 제j_k 실행 액션을 수행하기 이후에 획득된 상기 특정 오브젝트의 특정 최신 위치 정보, 상기 제j_k 실행 액션을 수행하기 이전에 획득된 상기 특정 오브젝트의 특정 최초 마스크 정보, 및 상기 제j_k 실행 액션을 수행하기 이후에 획득된 상기 특정 오브젝트의 특정 최신 마스크 정보 중 적어도 일부를 참조로 하여 결정됨 - 를 참조로 하여 상기 시맨틱 스페이셜 맵을 실시간으로 업데이트하는 것을 특징으로 하는 방법.

청구항 3

제2항에 있어서,

상기 (b) 단계에서,

상기 특정 오브젝트의 상기 특정 최신 마스크 정보를 상기 이미지 퍼셉션 모듈로부터 획득하고 상기 특정 최신 마스크 정보를 상기 환경 인지 메모리 모듈에 저장한 상태에서, 상기 제j_k 실행 액션의 제1 후속 실행 액션에

따른 타 오브젝트에 의해 상기 특정 오브젝트의 외형 중 일부가 차폐됨으로써 제2 후속 실행 액션에 따른 상기 특정 오브젝트의 특정 차폐 마스크 정보 - 상기 특정 차폐 마스크 정보는 상기 이미지 퍼셉션 모듈로부터 획득됨 - 와 상기 특정 최신 마스크 정보가 서로 불일치하는 것으로 판단되면, 상기 AI 에이전트가, 상기 특정 오브젝트의 상기 특정 최신 위치 정보를 파악한 다음, 상기 제2 후속 실행 액션에 따른 상기 특정 오브젝트의 상기 특정 차폐 마스크 정보를 상기 특정 최신 마스크 정보로 대체함으로써 상기 특정 오브젝트를 유효한 오브젝트로 판단하고, 상기 특정 오브젝트에 대한 상기 제2 후속 실행 액션을 수행하는 것을 특징으로 하는 방법.

청구항 4

제2항에 있어서,

상기 (b) 단계에서,

상기 특정 오브젝트와의 특정 인터랙션과, 상기 특정 오브젝트와 동일한 클래스를 가지는 적어도 하나의 별도 오브젝트와의 별도 인터랙션이 상기 자연어 지시 데이터에 포함되는 것으로 판단되면, 상기 AI 에이전트는, 상기 제j_k 실행 액션의 제3 후속 실행 액션에 따른 상기 별도 인터랙션을 수행하기 이전에, 상기 제j_k 실행 액션에 따른 상기 특정 오브젝트를 비탐색 타겟 오브젝트로서 결정하고, 상기 이미지 퍼셉션 모듈로부터 획득된 상기 특정 최신 위치 정보를 참조로 하여 상기 특정 오브젝트를 상기 제3 후속 실행 액션에 따른 상기 별도 오브젝트로서 무효한 오브젝트로 결정하는 것을 특징으로 하는 방법.

청구항 5

제2항에 있어서,

상기 (b) 단계에서,

상기 AI 에이전트가, 상기 이미지 퍼셉션 모듈로부터 상기 특정 오브젝트의 상기 특정 최신 마스크 정보를 획득하고, 상기 특정 최신 마스크 정보 및 상기 특정 오브젝트의 특정 최신 위치 정보를 상기 환경 인지 메모리 모듈에 저장하며, 상기 제j_k 실행 액션의 제4 후속 실행 액션에 따른 상기 특정 오브젝트에 대한 추가 인터랙션이 결정되면, 상기 특정 최신 위치 정보를 참조로 하여 상기 특정 오브젝트에 대한 최적의 탐색 경로를 결정하고, 상기 특정 최신 마스크 정보를 참조로 하여 상기 특정 오브젝트가 상기 제4 후속 실행 액션에 대해 유효한 오브젝트인지 확인하는 것을 특징으로 하는 방법.

청구항 6

제1항에 있어서,

상기 (a) 단계에서,

상기 AI 에이전트가, 상기 서브-골 플래너로 하여금, (i) 액션 홀더와, 상기 플레이스 홀더로 오브젝트 홀더 및 리셉터클 홀더를 포함하는 서브-골 프레임을 이용하여, 상기 액션 컴포넌트 시퀀스에 대응되는 제1 서브-골 액션 내지 제N 서브-골 액션 각각을 상기 서브-골 프레임의 상기 액션 홀더에 입력하여 제1 서브-골 프레임 내지 제N 서브-골 프레임을 생성하도록 하고, (ii) 상기 제1 서브-골 액션 내지 상기 제N 서브-골 액션에 따른 타겟 오브젝트와 타겟 리셉터클 사이의 관계 정보를 이용하여 상기 제1 서브-골 프레임 내지 상기 제N 서브-골 프레임 각각에서의 상기 오브젝트 홀더 및 상기 리셉터클 홀더에 대응되는 메타 클래스들을 생성함으로써 상기 서브-골 프레임 시퀀스를 생성하도록 하며, (iii) 상기 컨텍스트들 각각을 상기 서브-골 프레임 시퀀스에 대응되는 각각의 상기 메타 클래스들에 매칭하여 상기 제1 서브-골 내지 상기 제N 서브-골을 생성하도록 하는 것을 특징으로 하는 방법.

청구항 7

제6항에 있어서,

상기 (a) 단계에서,

상기 AI 에이전트가, 상기 서브-골 플래너로 하여금 하기 함수에 따라 상기 제1 서브-골 내지 상기 제N 서브-골을 생성하도록 하되,

$$f_{sub}(l) = \{S_j\}_{j=1}^{N}, \qquad S_j = (A_j, O_j, R_j)$$

상기 함수에서, f_{sub}는 상기 서브-골 플래너를 의미하고, l은 상기 자연어 지시 데이터를 의미하며, S_j는 전체 서브-골 중 j번째 서브-골을 의미하고, N은 서브-골의 전체 개수를 의미하며, A_j는 상기 j번째 서브-골에 대응하는 j번째 서브-골 액션을 의미하고, O_j는 상기 j번째 서브-골에 대응하는 j번째 타겟 오브젝트를 의미하며, R_j는 상기 j번째 서브-골에 대응하는 j번째 타겟 리셉터클을 의미하는 것을 특징으로 하는 방법.

청구항 8

제6항에 있어서,

상기 (a) 단계에서,

상기 AI 에이전트가, 상기 서브-골 플래너로 하여금 하기 함수에 따라 상기 컨텍스트를 예측하도록 하되,

$$f_{ctxt}^{O}(l) = c_O,\ f_{ctxt}^{M}(l) = c_M,\ f_{ctxt}^{R}(l) = c_R$$

상기 함수에서, l은 상기 자연어 지시 데이터를 의미하고, f_{ctxt}^{O}는 상기 태스크를 수행함에 있어 메인 타겟으로 결정되는 메인 오브젝트를 의미하며, f_{ctxt}^{M}는 상기 태스크를 수행함에 있어 상기 메인 오브젝트의 컨테이너로 결정되는 컨테이너 오브젝트를 의미하고, f_{ctxt}^{R}는 상기 태스크를 수행함에 있어 상기 메인 오브젝트 및 상기 컨테이너 오브젝트 중 적어도 하나의 목적지 타겟으로 결정되는 목적지 오브젝트를 의미하는 것을 특징으로 하는 방법.

청구항 9

제6항에 있어서,

상기 (a) 단계에서,

상기 AI 에이전트가, 상기 서브-골 플래너로 하여금 하기 함수에 따라 상기 서브-골 프레임을 생성하도록 하되,

$$f_{sf}(l) = \{F_j\}_{j=1}^{N},\ F_j = \left(A_j, \langle O \rangle_j, \langle R \rangle_j\right),\ \langle \cdot \rangle \in E \cup \{x_O, x_M, x_R\}$$

상기 함수에서, f_{sf}는 상기 서브-골 프레임을 의미하고, l은 상기 자연어 지시 데이터를 의미하며, F_j는 전체 서브-골 프레임 중 j번째 서브-골 프레임을 의미하고, N은 상기 서브-골 프레임의 전체 개수를 의미하며, A_j는 상기 j번째 서브-골 프레임에 대응하는 j번째 서브-골 액션을 의미하고, $\langle O \rangle_j$는 상기 j번째 서브-골 프레임에 대응하는 j번째 오브젝트 홀더를 의미하며, $\langle R \rangle_j$는 상기 j번째 서브-골 프레임에 대응하는 j번째 리셉터클 홀더를 의미하고, $\langle \cdot \rangle$는 상기 오브젝트 홀더 및 상기 리셉터클 홀더를 포함하는 전체 오브젝트 홀더를 의미하며, E는 상기 전체 환경 내의 전체 오브젝트의 집합을 의미하고, x_O는 상기 태스크를 수행함에 있어 메인 타겟으로 결정되는 메인 오브젝트가 매칭될 제1 메타클래스를 의미하며, x_M은 상기 메인 오브젝트의 컨테이너로 결정되는 컨테이너 오브젝트가 매칭될 제2 메타클래스를 의미하고, x_R은 상기 메인 오브젝트 및 상기 컨테이너 오브젝트 중 적어도 하나의 목적지 타겟으로 결정되는 목적지 오브젝트가 매칭될 제3 메타클래스를 의미하는 것을 특징으로 하는 방법.

청구항 10

제1항에 있어서,

상기 (a) 단계에서,

상기 AI 에이전트가, 상기 디테일드 플래너로 하여금 하기 함수에 따라 상기 제1 실행 액션 내지 상기 제T1 실

행 액션 및 상기 제1 실행 액션 내지 상기 제Tn 실행 액션을 생성하도록 하되,

$$f_{dp}^{g}\left((A_j, O_j, R_j)\right) = \{(a_k, o_k)\}_{k=1}^{Tj}$$

상기 함수에서, f_{dp}^{g}는 전체 서브-골 액션 중 j번째 서브-골에 대응하는 서브-골 액션 g - 상기 g 는 상기 j번째 서브-골에 대응하는 서브-골 액션인 A_j와 서로 동일함 - 에 대한 상기 디테일드 플래너를 의미하고, O_j는 상기 j번째 서브-골에 대응하는 j번째 타겟 오브젝트를 의미하며, R_j는 상기 j번째 서브-골에 대응하는 j번째 타겟 리셉터클을 의미하고, a_k는 상기 j번째 서브-골에 대한 k번째 실행 액션을 의미하며, o_k는 상기 j번째 서브-골에 대한 k번째 실행 오브젝트를 의미하고, 상기 Tj는 상기 j번째 서브-골에 대한 전체 실행 액션 개수를 의미하는 것을 특징으로 하는 방법.

청구항 11

컨텍스트 인지 플래닝 모듈 및 환경 인지 메모리 모듈을 포함하는 CAPEAM 모델에 따른 태스크를 수행하는 AI 에이전트에 있어서,

인스트럭션들을 저장하는 하나 이상의 메모리; 및

상기 인스트럭션들을 수행하도록 설정된 하나 이상의 프로세서를 포함하되, 상기 프로세서는, (I) 자연어 지시 데이터가 획득되면, 상기 자연어 지시 데이터를 서브-골 플래너로 입력하여 상기 서브-골 플래너로 하여금 상기 자연어 지시 데이터를 러닝 연산하도록 하여 상기 자연어 지시 데이터에 대응되는 서브-골 액션들의 액션 컴포넌트 시퀀스 및 상기 액션 컴포넌트 시퀀스에 대응되는 플레이스 홀더 컴포넌트 시퀀스를 포함하는 서브-골 프레임 시퀀스를 생성하도록 하고, 상기 자연어 지시 데이터로부터 획득한 컨텍스트들을 상기 컨텍스트들에 대응되는 상기 서브-골 프레임 시퀀스의 플레이스 홀더에 입력하여 상기 서브-골 프레임 시퀀스에 대응되는 제1 서브-골 내지 제N 서브-골 - 상기 N은 1 이상의 정수임- 을 생성하도록 하며, 디테일드 플래너로 하여금 상기 제1 서브-골에 대한 제1 실행 액션 내지 제T1 실행 액션 - 상기 T1은 1 이상의 정수임 -, 내지 상기 제N 서브-골에 대한 제1 실행 액션 내지 제Tn 실행 액션 - 상기 Tn은 1 이상의 정수임 - 을 생성하도록 하는 프로세스; 및 (II) 상기 제1 서브-골 내지 상기 제N 서브-골 중 어느 하나인 제j 서브-골에 대한 제k 실행 액션을 제j_k 실행 액션이라고 할 때, (i) 상기 j를 1부터 상기 N까지 증가시키며, 상기 k를 1부터 상기 Tj까지 증가시켜가며, 상기 제j_k 실행 액션에 대응하여, 이미지 퍼셉션 모듈을 통해 상기 제j_k 실행 액션에 따른 특정 오브젝트를 확인하고, (ii) 상기 태스크를 수행하는 전체 환경에 대한 시맨틱 스페이셜 맵 상에서의 상기 제j_k 실행 액션의 이전 실행 액션들에 대한 오브젝트들의 상태 정보를 저장하고 있는 상기 환경 인지 메모리 모듈을 참조하여 상기 특정 오브젝트가 유효한 오브젝트인지 확인하며, 상기 특정 오브젝트가 유효한 오브젝트로 확인되면 상기 특정 오브젝트에 대한 상기 제j_k 실행 액션을 수행하며, 상기 제j_k 실행 액션에 따른 상기 특정 오브젝트의 특정 상태 변화 정보를 상기 환경 인지 메모리 모듈에 저장하여 상기 시맨틱 스페이셜 맵을 업데이트하는 프로세스를 수행하는 것을 특징으로 하는 AI 에이전트.

청구항 12

제11항에 있어서,

상기 (II) 프로세스에서,

상기 프로세서가, 상기 이미지 퍼셉션 모듈로부터 획득된 상기 전체 환경에 대한 공간 정보를 참조로 하여 상기 전체 환경에 대응하는 뎁스-맵(depth map)을 생성하고, 상기 이미지 퍼셉션 모듈로부터 상기 전체 환경 내의 전체 오브젝트 중 적어도 일부 각각에 대응하는 전체 오브젝트 마스크 각각을 획득하며, 상기 전체 오브젝트 마스크 각각과 상기 뎁스-맵을 3D 세계 좌표로 백프로젝팅(back-projecting)하여 상기 시맨틱 스페이셜 맵을 구축하고, 상기 환경 인지 메모리 모듈에 저장된 상기 특정 상태 변화 정보 - 상기 특정 상태 변화 정보는, 상기 제j_k 실행 액션을 수행하기 이전에 획득된 상기 특정 오브젝트의 특정 최초 위치 정보, 상기 제j_k 실행 액션을 수행하기 이후에 획득된 상기 특정 오브젝트의 특정 최신 위치 정보, 상기 제j_k 실행 액션을 수행하기 이전에 획득된 상기 특정 오브젝트의 특정 최초 마스크 정보, 및 상기 제j_k 실행 액션을 수행하기 이후에 획득된 상기 특정 오브젝트의 특정 최신 마스크 정보 중 적어도 일부를 참조로 하여 결정됨 - 를 참조로 하여 상기 시맨틱 스페이셜 맵을 실시간으로 업데이트하는 것을 특징으로 하는 AI 에이전트.

청구항 13

제12항에 있어서,

상기 (II) 프로세스에서,

상기 특정 오브젝트의 상기 특정 최신 마스크 정보를 상기 이미지 퍼셉션 모듈로부터 획득하고 상기 특정 최신 마스크 정보를 상기 환경 인지 메모리 모듈에 저장한 상태에서, 상기 제j_k 실행 액션의 제1 후속 실행 액션에 따른 타 오브젝트에 의해 상기 특정 오브젝트의 외형 중 일부가 차폐됨으로써 제2 후속 실행 액션에 따른 상기 특정 오브젝트의 특정 차폐 마스크 정보 - 상기 특정 차폐 마스크 정보는 상기 이미지 퍼셉션 모듈로부터 획득됨 - 와 상기 특정 최신 마스크 정보가 서로 불일치하는 것으로 판단되면, 상기 프로세서가, 상기 특정 오브젝트의 상기 특정 최신 위치 정보를 파악한 다음, 상기 제2 후속 실행 액션에 따른 상기 특정 오브젝트의 상기 특정 차폐 마스크 정보를 상기 특정 최신 마스크 정보로 대체함으로써 상기 특정 오브젝트를 유효한 오브젝트로 판단하고, 상기 특정 오브젝트에 대한 상기 제2 후속 실행 액션을 수행하는 것을 특징으로 하는 AI 에이전트.

청구항 14

제12항에 있어서,

상기 (II) 프로세스에서,

상기 특정 오브젝트와의 특정 인터랙션과, 상기 특정 오브젝트와 동일한 클래스를 가지는 적어도 하나의 별도 오브젝트와의 별도 인터랙션이 상기 자연어 지시 데이터에 포함되는 것으로 판단되면, 상기 프로세서가, 상기 제j_k 실행 액션의 제3 후속 실행 액션에 따른 상기 별도 인터랙션을 수행하기 이전에, 상기 제j_k 실행 액션에 따른 상기 특정 오브젝트를 비탐색 타겟 오브젝트로서 결정하고, 상기 이미지 퍼셉션 모듈로부터 획득된 상기 특정 최신 위치 정보를 참조로 하여 상기 특정 오브젝트를 상기 제3 후속 실행 액션에 따른 상기 별도 오브젝트로서 무효한 오브젝트로 결정하는 것을 특징으로 하는 AI 에이전트.

청구항 15

제12항에 있어서,

상기 (II) 프로세스에서,

상기 프로세서가, 상기 이미지 퍼셉션 모듈로부터 상기 특정 오브젝트의 상기 특정 최신 마스크 정보를 획득하고, 상기 특정 최신 마스크 정보 및 상기 특정 오브젝트의 특정 최신 위치 정보를 상기 환경 인지 메모리 모듈에 저장하며, 상기 제j_k 실행 액션의 제4 후속 실행 액션에 따른 상기 특정 오브젝트에 대한 추가 인터랙션이 결정되면, 상기 특정 최신 위치 정보를 참조로 하여 상기 특정 오브젝트에 대한 최적의 탐색 경로를 결정하고, 상기 특정 최신 마스크 정보를 참조로 하여 상기 특정 오브젝트가 상기 제4 후속 실행 액션에 대해 유효한 오브젝트인지 확인하는 것을 특징으로 하는 AI 에이전트.

청구항 16

제11항에 있어서,

상기 (I) 프로세스에서,

상기 프로세서가, 상기 서브-골 플래너로 하여금, (i) 액션 홀더와, 상기 플레이스 홀더로 오브젝트 홀더 및 리셉터클 홀더를 포함하는 서브-골 프레임을 이용하여, 상기 액션 컴포넌트 시퀀스에 대응되는 제1 서브-골 액션 내지 제N 서브-골 액션 각각을 상기 서브-골 프레임의 상기 액션 홀더에 입력하여 제1 서브-골 프레임 내지 제N 서브-골 프레임을 생성하도록 하고, (ii) 상기 제1 서브-골 액션 내지 상기 제N 서브-골 액션에 따른 타겟 오브젝트와 타겟 리셉터클 사이의 관계 정보를 이용하여 상기 제1 서브-골 프레임 내지 상기 제N 서브-골 프레임 각각에서의 상기 오브젝트 홀더 및 상기 리셉터클 홀더에 대응되는 메타 클래스들을 생성함으로써 상기 서브-골 프레임 시퀀스를 생성하도록 하며, (iii) 상기 컨텍스트들 각각을 상기 서브-골 프레임 시퀀스에 대응되는 각각의 상기 메타 클래스들에 매칭하여 상기 제1 서브-골 내지 상기 제N 서브-골을 생성하도록 하는 것을 특징으로 하는 AI 에이전트.

청구항 17

제16항에 있어서,

상기 (I) 프로세스에서,

상기 프로세서가, 상기 서브-골 플래너로 하여금 하기 함수에 따라 상기 제1 서브-골 내지 상기 제N 서브-골을 생성하도록 하되,

$$f_{sub}(l) = \{S_j\}_{j=1}^{N}, \quad S_j = (A_j, O_j, R_j)$$

상기 함수에서, f_{sub}는 상기 서브-골 플래너를 의미하고, l은 상기 자연어 지시 데이터를 의미하며, S_j는 전체 서브-골 중 j번째 서브-골을 의미하고, N은 서브-골의 전체 개수를 의미하며, A_j는 상기 j번째 서브-골에 대응하는 j번째 서브-골 액션을 의미하고, O_j는 상기 j번째 서브-골에 대응하는 j번째 타겟 오브젝트를 의미하며, R_j는 상기 j번째 서브-골에 대응하는 j번째 타겟 리셉터클을 의미하는 것을 특징으로 하는 AI 에이전트.

청구항 18

제16항에 있어서,

상기 (I) 프로세스에서,

상기 프로세서가, 상기 서브-골 플래너로 하여금 하기 함수에 따라 상기 컨텍스트를 예측하도록 하되,

$$f_{ctxt}^{O}(l) = c_O, \quad f_{ctxt}^{M}(l) = c_M, \quad f_{ctxt}^{R}(l) = c_R$$

상기 함수에서, l은 상기 자연어 지시 데이터를 의미하고, f_{ctxt}^{O}는 상기 태스크를 수행함에 있어 메인 타겟으로 결정되는 메인 오브젝트를 의미하며, f_{ctxt}^{M}는 상기 태스크를 수행함에 있어 상기 메인 오브젝트의 컨테이너로 결정되는 컨테이너 오브젝트를 의미하고, f_{ctxt}^{R}는 상기 태스크를 수행함에 있어 상기 메인 오브젝트 및 상기 컨테이너 오브젝트 중 적어도 하나의 목적지 타겟으로 결정되는 목적지 오브젝트를 의미하는 것을 특징으로 하는 AI 에이전트.

청구항 19

제16항에 있어서,

상기 (I) 프로세스에서,

상기 프로세서가, 상기 서브-골 플래너로 하여금 하기 함수에 따라 상기 서브-골 프레임을 생성하도록 하되,

$$f_{sf}(l) = \{F_j\}_{j=1}^{N}, \quad F_j = \left(A_j, \langle O \rangle_j, \langle R \rangle_j\right), \quad \langle \cdot \rangle \in \mathrm{E} \cup \{x_O, x_M, x_R\}$$

상기 함수에서, f_{sf}는 상기 서브-골 프레임을 의미하고, l은 상기 자연어 지시 데이터를 의미하며, F_j는 전체 서브-골 프레임 중 j번째 서브-골 프레임을 의미하고, N은 상기 서브-골 프레임의 전체 개수를 의미하며, A_j는 상기 j번째 서브-골 프레임에 대응하는 j번째 서브-골 액션을 의미하고, $\langle O \rangle_j$는 상기 j번째 서브-골 프레임에 대응하는 j번째 오브젝트 홀더를 의미하며, $\langle R \rangle_j$는 상기 j번째 서브-골 프레임에 대응하는 j번째 리셉터클 홀더를 의미하고, $\langle \cdot \rangle$는 상기 오브젝트 홀더 및 상기 리셉터클 홀더를 포함하는 전체 오브젝트 홀더를 의미하며, E는 상기 전체 환경 내의 전체 오브젝트의 집합을 의미하고, x_O는 상기 태스크를 수행함에 있어 메인 타겟으로 결정되는 메인 오브젝트가 매칭될 제1 메타클래스를 의미하며, x_M은 상기 메인 오브젝트의 컨테이너로 결정되는 컨테이너 오브젝트가 매칭될 제2 메타클래스를 의미하고, x_R은 상기 메인 오브젝트 및 상기 컨테이너 오브젝트 중 적어도 하나의 목적지 타겟으로 결정되는 목적지 오브젝트가 매칭될 제3 메타클래스를 의미하는 것을 특징으

로 하는 AI 에이전트.

청구항 20

제11항에 있어서,

상기 (I) 프로세스에서,

상기 프로세서가, 상기 디테일드 플래너로 하여금 하기 함수에 따라 상기 제1 실행 액션 내지 상기 제T1 실행 액션 및 상기 제1 실행 액션 내지 상기 제Tn 실행 액션을 생성하도록 하되,

$$f_{dp}^{g}\left((A_j, O_j, R_j)\right) = \{(a_k, o_k)\}_{k=1}^{T_j}$$

상기 함수에서, f_{dp}^{g}는 전체 서브-골 액션 중 j번째 서브-골에 대응하는 서브-골 액션 g - 상기 g 는 상기 j번째 서브-골에 대응하는 서브-골 액션인 A_j와 서로 동일함 - 에 대한 상기 디테일드 플래너를 의미하고, O_j는 상기 j번째 서브-골에 대응하는 j번째 타겟 오브젝트를 의미하며, R_j는 상기 j번째 서브-골에 대응하는 j번째 타겟 리셉터클을 의미하고, a_k는 상기 j번째 서브-골에 대한 k번째 실행 액션을 의미하며, o_k는 상기 j번째 서브-골에 대한 k번째 실행 오브젝트를 의미하고, 상기 T_j는 상기 j번째 서브-골에 대한 전체 실행 액션 개수를 의미하는 것을 특징으로 하는 AI 에이전트.

발명의 설명

기 술 분 야

[0001] 본 발명은 컨텍스트 인지 플래닝(Context-Aware Planning) 모듈 및 환경 인지 메모리(Environment-Aware Memory) 모듈을 포함하는 CAPEAM 모델에 따른 AI 에이전트의 태스크 수행 방법 및 이를 사용한 AI 에이전트에 관한 것이다.

배 경 기 술

[0002] 언어 지시에 따라 집안일 등의 귀찮은 일들을 수행할 수 있는 AI 부하를 갖는 것은 모두가 꿈꾸는 일이다. AI가 이와 같은 일들을 대신 수행할 수 있으려면, AI는 시각적으로 풍부한 3D 환경에서 탐색, 오브젝트 인터랙션 및 대화형 추론을 할 수 있어야 한다. 한 걸음 더 나아가, AI가 자기중심적 비전에 기반하여 자연어 지시에 따라 환경을 탐색하고, 객체와 인터랙션하며, 장기적인 업무를 해낼 수 있다면 더욱 이상적일 것이다.

[0003] 한편, 종래에는 AI가 자연어 지시에 따라 환경을 탐색하고 오브젝트와의 인터랙션을 시도하는 기술이 있었으나, AI가 장기적인 일련의 작업을 수행하는 과정에서 올바른 오브젝트가 무엇인지 정확히 인지하지 못하여 작업과 무관한 오브젝트와의 인터랙션을 시도하는 문제점이 있었다.

[0004] 또한, 종래에는 AI가 오브젝트의 상태 변화와 오브젝트의 위치 변화를 추적하지 못하여, 상태가 변화된 오브젝트 및 위치가 변경된 오브젝트를 정확히 탐색하지 못하는 문제점이 있었다.

[0005] 따라서, 이러한 문제점들을 해결하기 위한 개선 방안이 요구되는 실정이다.

발명의 내용

해결하려는 과제

[0006] 본 발명은 상술한 문제점을 모두 해결하는 것을 그 목적으로 한다.

[0007] 또한, 본 발명은 서브-골 플래너로 하여금 획득된 자연어 지시 데이터를 러닝 연산하도록 하여 자연어 지시 데이터에 대응되는 서브-골 액션들의 액션 컴포넌트 시퀀스 및 액션 컴포넌트 시퀀스에 대응되는 플레이스 홀더 컴포넌트 시퀀스를 포함하는 서브-골 프레임 시퀀스를 생성하도록 하고, 자연어 지시 데이터로부터 획득한 컨텍스트들을 컨텍스트들에 대응되는 서브-골 프레임 시퀀스의 플레이스 홀더에 입력하여 서브-골 프레임 시퀀스에 대응되는 제1 서브-골 내지 제N 서브-골을 생성하도록 하며, 디테일드 플래너로 하여금 제1 서브-골에 대한 제1 실행 액션 내지 제T1 실행 액션, 내지 제N 서브-골에 대한 제1 실행 액션 내지 제Tn 실행 액션을 생성하도록 하

는 것을 다른 목적으로 한다.

[0008] 또한, 본 발명은 제1 서브-골 내지 제N 서브-골 중 어느 하나인 제j_k 실행 액션에 대응하여, 이미지 퍼셉션 모듈을 통해 제j_k 실행 액션에 따른 특정 오브젝트를 확인하고, 태스크를 수행하는 전체 환경에 대한 시맨틱 스페이셜 맵 상에서의 제j_k 실행 액션의 이전 실행 액션들에 대한 오브젝트들의 상태 정보를 저장하고 있는 환경 인지 메모리 모듈을 참조하여 특정 오브젝트가 유효한 오브젝트인지 확인하며, 특정 오브젝트가 유효한 오브젝트로 확인되면 특정 오브젝트에 대한 제j_k 실행 액션을 수행하며, 제j_k 실행 액션에 따른 특정 오브젝트의 특정 상태 변화 정보를 환경 인지 메모리 모듈에 저장하여 시맨틱 스페이셜 맵을 업데이트하는 것을 또 다른 목적으로 한다.

과제의 해결 수단

[0009] 상기한 바와 같은 본 발명의 목적을 달성하고, 후술하는 본 발명의 특징적인 효과를 실현하기 위한, 본 발명의 특징적인 구성은 하기와 같다.

[0010] 본 발명의 일 태양에 따르면, 컨텍스트 인지 플래닝 모듈 및 환경 인지 메모리 모듈을 포함하는 CAPEAM 모델에 따른 태스크 수행 방법에 있어서, (a) 자연어 지시 데이터가 획득되면, AI 에이전트가, 상기 자연어 지시 데이터를 서브-골 플래너로 입력하여 상기 서브-골 플래너로 하여금 상기 자연어 지시 데이터를 러닝 연산하도록 하여 상기 자연어 지시 데이터에 대응되는 서브-골 액션들의 액션 컴포넌트 시퀀스 및 상기 액션 컴포넌트 시퀀스에 대응되는 플레이스 홀더 컴포넌트 시퀀스를 포함하는 서브-골 프레임 시퀀스를 생성하도록 하고, 상기 자연어 지시 데이터로부터 획득한 컨텍스트들을 상기 컨텍스트들에 대응되는 상기 서브-골 프레임 시퀀스의 플레이스 홀더에 입력하여 상기 서브-골 프레임 시퀀스에 대응되는 제1 서브-골 내지 제N 서브-골 - 상기 N은 1 이상의 정수임- 을 생성하도록 하며, 디테일드 플래너로 하여금 상기 제1 서브-골에 대한 제1 실행 액션 내지 제T1 실행 액션 - 상기 T1은 1 이상의 정수임 -, 내지 상기 제N 서브-골에 대한 제1 실행 액션 내지 제Tn 실행 액션 - 상기 Tn은 1 이상의 정수임 - 을 생성하도록 하는 단계; 및 (b) 상기 제1 서브-골 내지 상기 제N 서브-골 중 어느 하나인 제j 서브-골에 대한 제k 실행 액션을 제j_k 실행 액션이라고 할 때, 상기 AI 에이전트가, (i) 상기 j를 1부터 상기 N까지 증가시키며, 상기 k를 1부터 상기 Tj까지 증가시켜가며, 상기 제j_k 실행 액션에 대응하여, 이미지 퍼셉션 모듈을 통해 상기 제j_k 실행 액션에 따른 특정 오브젝트를 확인하고, (ii) 상기 태스크를 수행하는 전체 환경에 대한 시맨틱 스페이셜 맵 상에서의 상기 제j_k 실행 액션의 이전 실행 액션들에 대한 오브젝트들의 상태 정보를 저장하고 있는 상기 환경 인지 메모리 모듈을 참조하여 상기 특정 오브젝트가 유효한 오브젝트인지 확인하며, 상기 특정 오브젝트가 유효한 오브젝트로 확인되면 상기 특정 오브젝트에 대한 상기 제j_k 실행 액션을 수행하며, 상기 제j_k 실행 액션에 따른 상기 특정 오브젝트의 특정 상태 변화 정보를 상기 환경 인지 메모리 모듈에 저장하여 상기 시맨틱 스페이셜 맵을 업데이트하는 단계;를 포함하는 방법이 개시된다.

[0011] 일례로서, 상기 (b) 단계에서, 상기 AI 에이전트가, 상기 이미지 퍼셉션 모듈로부터 획득된 상기 전체 환경에 대한 공간 정보를 참조로 하여 상기 전체 환경에 대응하는 뎁스-맵(depth map)을 생성하고, 상기 이미지 퍼셉션 모듈로부터 상기 전체 환경 내의 전체 오브젝트 중 적어도 일부 각각에 대응하는 전체 오브젝트 마스크 각각을 획득하며, 상기 전체 오브젝트 마스크 각각과 상기 뎁스-맵을 3D 세계 좌표로 백프로젝팅(back-projecting)하여 상기 시맨틱 스페이셜 맵을 구축하고, 상기 환경 인지 메모리 모듈에 저장된 상기 특정 상태 변화 정보 - 상기 특정 상태 변화 정보는, 상기 제j_k 실행 액션을 수행하기 이전에 획득된 상기 특정 오브젝트의 특정 최초 위치 정보, 상기 제j_k 실행 액션을 수행하기 이후에 획득된 상기 특정 오브젝트의 특정 최신 위치 정보, 상기 제j_k 실행 액션을 수행하기 이전에 획득된 상기 특정 오브젝트의 특정 최초 마스크 정보, 및 상기 제j_k 실행 액션을 수행하기 이후에 획득된 상기 특정 오브젝트의 특정 최신 마스크 정보 중 적어도 일부를 참조로 하여 결정됨 - 를 참조로 하여 상기 시맨틱 스페이셜 맵을 실시간으로 업데이트하는 것을 특징으로 하는 방법이 개시된다.

[0012] 일례로서, 상기 (b) 단계에서, 상기 특정 오브젝트의 상기 특정 최신 마스크 정보를 상기 이미지 퍼셉션 모듈로부터 획득하고 상기 특정 최신 마스크 정보를 상기 환경 인지 메모리 모듈에 저장한 상태에서, 상기 제j_k 실행 액션의 제1 후속 실행 액션에 따른 타 오브젝트에 의해 상기 특정 오브젝트의 외형 중 일부가 차폐됨으로써 제2 후속 실행 액션에 따른 상기 특정 오브젝트의 특정 차폐 마스크 정보 - 상기 특정 차폐 마스크 정보는 상기 이미지 퍼셉션 모듈로부터 획득됨 - 와 상기 특정 최신 마스크 정보가 서로 불일치하는 것으로 판단되면, 상기 AI 에이전트가, 상기 특정 오브젝트의 상기 특정 최신 위치 정보를 파악한 다음, 상기 제2 후속 실행 액션에 따른 상기 특정 오브젝트의 상기 특정 차폐 마스크 정보를 상기 특정 최신 마스크 정보로 대체함으로써 상기 특정 오브젝트를 유효한 오브젝트로 판단하고, 상기 특정 오브젝트에 대한 상기 제2 후속 실행 액션을 수행하는 것을 특징으로 하는 방법이 개시된다.

[0013] 일례로서, 상기 (b) 단계에서, 상기 특정 오브젝트와의 특정 인터랙션과, 상기 특정 오브젝트와 동일한 클래스를 가지는 적어도 하나의 별도 오브젝트와의 별도 인터랙션이 상기 자연어 지시 데이터에 포함되는 것으로 판단되면, 상기 AI 에이전트는, 상기 제j_k 실행 액션의 제3 후속 실행 액션에 따른 상기 별도 인터랙션을 수행하기 이전에, 상기 제j_k 실행 액션에 따른 상기 특정 오브젝트를 비탐색 타겟 오브젝트로서 결정하고, 상기 이미지 퍼셉션 모듈로부터 획득된 상기 특정 최신 위치 정보를 참조로 하여 상기 특정 오브젝트를 상기 제3 후속 실행 액션에 따른 상기 별도 오브젝트로서 무효한 오브젝트로 결정하는 것을 특징으로 하는 방법이 개시된다.

[0014] 일례로서, 상기 (b) 단계에서, 상기 AI 에이전트가, 상기 이미지 퍼셉션 모듈로부터 상기 특정 오브젝트의 상기 특정 최신 마스크 정보를 획득하고, 상기 특정 최신 마스크 정보 및 상기 특정 오브젝트의 특정 최신 위치 정보를 상기 환경 인지 메모리 모듈에 저장하며, 상기 제j_k 실행 액션의 제4 후속 실행 액션에 따른 상기 특정 오브젝트에 대한 추가 인터랙션이 결정되면, 상기 특정 최신 위치 정보를 참조로 하여 상기 특정 오브젝트에 대한 최적의 탐색 경로를 결정하고, 상기 특정 최신 마스크 정보를 참조로 하여 상기 특정 오브젝트가 상기 제4 후속 실행 액션에 대해 유효한 오브젝트인지 확인하는 것을 특징으로 하는 방법이 개시된다.

[0015] 일례로서, 상기 (a) 단계에서, 상기 AI 에이전트가, 상기 서브-골 플래너로 하여금, (i) 액션 홀더와, 상기 플레이스 홀더로 오브젝트 홀더 및 리셉터클 홀더를 포함하는 서브-골 프레임을 이용하여, 상기 액션 컴포넌트 시퀀스에 대응되는 제1 서브-골 액션 내지 제N 서브-골 액션 각각을 상기 서브-골 프레임의 상기 액션 홀더에 입력하여 제1 서브-골 프레임 내지 제N 서브-골 프레임을 생성하도록 하고, (ii) 상기 제1 서브-골 액션 내지 상기 제N 서브-골 액션에 따른 타겟 오브젝트와 타겟 리셉터클 사이의 관계 정보를 이용하여 상기 제1 서브-골 프레임 내지 상기 제N 서브-골 프레임 각각에서의 상기 오브젝트 홀더 및 상기 리셉터클 홀더에 대응되는 메타 클래스들을 생성함으로써 상기 서브-골 프레임 시퀀스를 생성하도록 하며, (iii) 상기 컨텍스트들 각각을 상기 서브-골 프레임 시퀀스에 대응되는 각각의 상기 메타 클래스들에 매칭하여 상기 제1 서브-골 내지 상기 제N 서브-골을 생성하도록 하는 것을 특징으로 하는 방법이 개시된다.

[0016] 일례로서, 상기 (a) 단계에서, 상기 AI 에이전트가, 상기 서브-골 플래너로 하여금 하기 함수에 따라 상기 제1 서브-골 내지 상기 제N 서브-골을 생성하도록 하되,

[0017] $$f_{sub}(l) = \{S_j\}_{j=1}^{N}, \qquad S_j = (A_j, O_j, R_j)$$

[0018] 상기 함수에서, f_{sub} 는 상기 서브-골 플래너를 의미하고, l 은 상기 자연어 지시 데이터를 의미하며, S_j 는 전체 서브-골 중 j번째 서브-골을 의미하고, N 은 서브-골의 전체 개수를 의미하며, A_j 는 상기 j번째 서브-골에 대응하는 j번째 서브-골 액션을 의미하고, O_j 는 상기 j번째 서브-골에 대응하는 j번째 타겟 오브젝트를 의미하며, R_j 는 상기 j번째 서브-골에 대응하는 j번째 타겟 리셉터클을 의미하는 것을 특징으로 하는 방법이 개시된다.

[0019] 일례로서, 상기 (a) 단계에서, 상기 AI 에이전트가, 상기 서브-골 플래너로 하여금 하기 함수에 따라 상기 컨텍스트를 예측하도록 하되,

[0020] $$f_{ctxt}^{O}(l) = c_O, \quad f_{ctxt}^{M}(l) = c_M, \quad f_{ctxt}^{R}(l) = c_R$$

[0021] 상기 함수에서, l 은 상기 자연어 지시 데이터를 의미하고, f_{ctxt}^{O} 는 상기 태스크를 수행함에 있어 메인 타겟으로 결정되는 메인 오브젝트를 의미하며, f_{ctxt}^{M} 는 상기 태스크를 수행함에 있어 상기 메인 오브젝트의 컨테이너로 결정되는 컨테이너 오브젝트를 의미하고, f_{ctxt}^{R} 는 상기 태스크를 수행함에 있어 상기 메인 오브젝트 및 상기 컨테이너 오브젝트 중 적어도 하나의 목적지 타겟으로 결정되는 목적지 오브젝트를 의미하는 것을 특징으로 하는 방법이 개시된다.

[0022] 일례로서, 상기 (a) 단계에서, 상기 AI 에이전트가, 상기 서브-골 플래너로 하여금 하기 함수에 따라 상기 서브-골 프레임을 생성하도록 하되,

[0023] $$f_{sf}(l) = \{F_j\}_{j=1}^{N}, \quad F_j = \left(A_j, \langle O \rangle_j, \langle R \rangle_j\right), \quad \langle \cdot \rangle \in \mathrm{E} \cup \{x_O, x_M, x_R\}$$

[0024] 상기 함수에서, f_{sf} 는 상기 서브-골 프레임을 의미하고, l 은 상기 자연어 지시 데이터를 의미하며, F_j 는 전체

서브-골 프레임 중 j번째 서브-골 프레임을 의미하고, N은 상기 서브-골 프레임의 전체 개수를 의미하며, A_j는 상기 j번째 서브-골 프레임에 대응하는 j번째 서브-골 액션을 의미하고, $\langle O \rangle_j$는 상기 j번째 서브-골 프레임에 대응하는 j번째 오브젝트 홀더를 의미하며, $\langle R \rangle_j$는 상기 j번째 서브-골 프레임에 대응하는 j번째 리셉터클 홀더를 의미하고, $\langle \cdot \rangle$는 상기 오브젝트 홀더 및 상기 리셉터클 홀더를 포함하는 전체 오브젝트 홀더를 의미하며, E는 상기 전체 환경 내의 전체 오브젝트의 집합을 의미하고, x_O는 상기 태스크를 수행함에 있어 메인 타겟으로 결정되는 메인 오브젝트가 매칭될 제1 메타클래스를 의미하며, x_M은 상기 메인 오브젝트의 컨테이너로 결정되는 컨테이너 오브젝트가 매칭될 제2 메타클래스를 의미하고, x_R은 상기 메인 오브젝트 및 상기 컨테이너 오브젝트 중 적어도 하나의 목적지 타겟으로 결정되는 목적지 오브젝트가 매칭될 제3 메타클래스를 의미하는 것을 특징으로 하는 방법이 개시된다.

[0025] 일례로서, 상기 (a) 단계에서, 상기 AI 에이전트가, 상기 디테일드 플래너로 하여금 하기 함수에 따라 상기 제1 실행 액션 내지 상기 제T1 실행 액션 및 상기 제1 실행 액션 내지 상기 제Tn 실행 액션을 생성하도록 하되,

[0026] $$f_{dp}^{g}\left((A_j, O_j, R_j)\right) = \{(a_k, o_k)\}_{k=1}^{Tj}$$

[0027] 상기 함수에서, f_{dp}^{g}는 전체 서브-골 액션 중 j번째 서브-골에 대응하는 서브-골 액션 ^{g} - 상기 ^{g}는 상기 j번째 서브-골에 대응하는 서브-골 액션인 A_j와 서로 동일함 - 에 대한 상기 디테일드 플래너를 의미하고, O_j는 상기 j번째 서브-골에 대응하는 j번째 타겟 오브젝트를 의미하며, R_j는 상기 j번째 서브-골에 대응하는 j번째 타겟 리셉터클을 의미하고, a_k는 상기 j번째 서브-골에 대한 k번째 실행 액션을 의미하며, o_k는 상기 j번째 서브-골에 대한 k번째 실행 오브젝트를 의미하고, 상기 T_j는 상기 j번째 서브-골에 대한 전체 실행 액션 개수를 의미하는 것을 특징으로 하는 방법이 개시된다.

[0028] 본 발명의 또 다른 태양에 따르면, 컨텍스트 인지 플래닝 모듈 및 환경 인지 메모리 모듈을 포함하는 CAPEAM 모델에 따른 태스크를 수행하는 AI 에이전트에 있어서, 인스트럭션들을 저장하는 하나 이상의 메모리; 및 상기 인스트럭션들을 수행하도록 설정된 하나 이상의 프로세서를 포함하되, 상기 프로세서는, (I) 자연어 지시 데이터가 획득되면, 상기 자연어 지시 데이터를 서브-골 플래너로 입력하여 상기 서브-골 플래너로 하여금 상기 자연어 지시 데이터를 러닝 연산하도록 하여 상기 자연어 지시 데이터에 대응되는 서브-골 액션들의 액션 컴포넌트 시퀀스 및 상기 액션 컴포넌트 시퀀스에 대응되는 플레이스 홀더 컴포넌트 시퀀스를 포함하는 서브-골 프레임 시퀀스를 생성하도록 하고, 상기 자연어 지시 데이터로부터 획득한 컨텍스트들을 상기 컨텍스트들에 대응되는 상기 서브-골 프레임 시퀀스의 플레이스 홀더에 입력하여 상기 서브-골 프레임 시퀀스에 대응되는 제1 서브-골 내지 제N 서브-골 - 상기 N은 1 이상의 정수임- 을 생성하도록 하며, 디테일드 플래너로 하여금 상기 제1 서브-골에 대한 제1 실행 액션 내지 제T1 실행 액션 - 상기 T1은 1 이상의 정수임 -, 내지 상기 제N 서브-골에 대한 제1 실행 액션 내지 제Tn 실행 액션 - 상기 Tn은 1 이상의 정수임 - 을 생성하도록 하는 프로세스; 및 (II) 상기 제1 서브-골 내지 상기 제N 서브-골 중 어느 하나인 제j 서브-골에 대한 제k 실행 액션을 제j_k 실행 액션이라고 할 때, (i) 상기 j를 1부터 상기 N까지 증가시키며, 상기 k를 1부터 상기 Tj까지 증가시켜가며, 상기 제j_k 실행 액션에 대응하여, 이미지 퍼셉션 모듈을 통해 상기 제j_k 실행 액션에 따른 특정 오브젝트를 확인하고, (ii) 상기 태스크를 수행하는 전체 환경에 대한 시맨틱 스페이셜 맵 상에서의 상기 제j_k 실행 액션의 이전 실행 액션들에 대한 오브젝트들의 상태 정보를 저장하고 있는 상기 환경 인지 메모리 모듈을 참조하여 상기 특정 오브젝트가 유효한 오브젝트인지 확인하며, 상기 특정 오브젝트가 유효한 오브젝트로 확인되면 상기 특정 오브젝트에 대한 상기 제j_k 실행 액션을 수행하며, 상기 제j_k 실행 액션에 따른 상기 특정 오브젝트의 특정 상태 변화 정보를 상기 환경 인지 메모리 모듈에 저장하여 상기 시맨틱 스페이셜 맵을 업데이트하는 프로세스를 수행하는 것을 특징으로 하는 AI 에이전트가 개시된다.

[0029] 일례로서, 상기 (II) 프로세스에서, 상기 프로세서가, 상기 이미지 퍼셉션 모듈로부터 획득된 상기 전체 환경에

대한 공간 정보를 참조로 하여 상기 전체 환경에 대응하는 뎁스-맵(depth map)을 생성하고, 상기 이미지 퍼셉션 모듈로부터 상기 전체 환경 내의 전체 오브젝트 중 적어도 일부 각각에 대응하는 전체 오브젝트 마스크 각각을 획득하며, 상기 전체 오브젝트 마스크 각각과 상기 뎁스-맵을 3D 세계 좌표로 백프로젝팅(back-projecting)하여 상기 시맨틱 스페이셜 맵을 구축하고, 상기 환경 인지 메모리 모듈에 저장된 상기 특정 상태 변화 정보 - 상기 특정 상태 변화 정보는, 상기 제j_k 실행 액션을 수행하기 이전에 획득된 상기 특정 오브젝트의 특정 최초 위치 정보, 상기 제j_k 실행 액션을 수행하기 이후에 획득된 상기 특정 오브젝트의 특정 최신 위치 정보, 상기 제j_k 실행 액션을 수행하기 이전에 획득된 상기 특정 오브젝트의 특정 최초 마스크 정보, 및 상기 제j_k 실행 액션을 수행하기 이후에 획득된 상기 특정 오브젝트의 특정 최신 마스크 정보 중 적어도 일부를 참조로 하여 결정됨 - 를 참조로 하여 상기 시맨틱 스페이셜 맵을 실시간으로 업데이트하는 것을 특징으로 하는 AI 에이전트가 개시된다.

[0030] 일례로서, 상기 (II) 프로세스에서, 상기 특정 오브젝트의 상기 특정 최신 마스크 정보를 상기 이미지 퍼셉션 모듈로부터 획득하고 상기 특정 최신 마스크 정보를 상기 환경 인지 메모리 모듈에 저장한 상태에서, 상기 제j_k 실행 액션의 제1 후속 실행 액션에 따른 타 오브젝트에 의해 상기 특정 오브젝트의 외형 중 일부가 차폐됨으로써 제2 후속 실행 액션에 따른 상기 특정 오브젝트의 특정 차폐 마스크 정보 - 상기 특정 차폐 마스크 정보는 상기 이미지 퍼셉션 모듈로부터 획득됨 - 와 상기 특정 최신 마스크 정보가 서로 불일치하는 것으로 판단되면, 상기 프로세서가, 상기 특정 오브젝트의 상기 특정 최신 위치 정보를 파악한 다음, 상기 제2 후속 실행 액션에 따른 상기 특정 오브젝트의 상기 특정 차폐 마스크 정보를 상기 특정 최신 마스크 정보로 대체함으로써 상기 특정 오브젝트를 유효한 오브젝트로 판단하고, 상기 특정 오브젝트에 대한 상기 제2 후속 실행 액션을 수행하는 것을 특징으로 하는 AI 에이전트가 개시된다.

[0031] 일례로서, 상기 (II) 프로세스에서, 상기 특정 오브젝트와의 특정 인터랙션과, 상기 특정 오브젝트와 동일한 클래스를 가지는 적어도 하나의 별도 오브젝트와의 별도 인터랙션이 상기 자연어 지시 데이터에 포함되는 것으로 판단되면, 상기 프로세서가, 상기 제j_k 실행 액션의 제3 후속 실행 액션에 따른 상기 별도 인터랙션을 수행하기 이전에, 상기 제j_k 실행 액션에 따른 상기 특정 오브젝트를 비탐색 타겟 오브젝트로서 결정하고, 상기 이미지 퍼셉션 모듈로부터 획득된 상기 특정 최신 위치 정보를 참조로 하여 상기 특정 오브젝트를 상기 제3 후속 실행 액션에 따른 상기 별도 오브젝트로서 무효한 오브젝트로 결정하는 것을 특징으로 하는 AI 에이전트가 개시된다.

[0032] 일례로서, 상기 (II) 프로세스에서, 상기 프로세서가, 상기 이미지 퍼셉션 모듈로부터 상기 특정 오브젝트의 상기 특정 최신 마스크 정보를 획득하고, 상기 특정 최신 마스크 정보 및 상기 특정 오브젝트의 특정 최신 위치 정보를 상기 환경 인지 메모리 모듈에 저장하며, 상기 제j_k 실행 액션의 제4 후속 실행 액션에 따른 상기 특정 오브젝트에 대한 추가 인터랙션이 결정되면, 상기 특정 최신 위치 정보를 참조로 하여 상기 특정 오브젝트에 대한 최적의 탐색 경로를 결정하고, 상기 특정 최신 마스크 정보를 참조로 하여 상기 특정 오브젝트가 상기 제4 후속 실행 액션에 대해 유효한 오브젝트인지 확인하는 것을 특징으로 하는 AI 에이전트가 개시된다.

[0033] 일례로서, 상기 (I) 프로세스에서, 상기 프로세서가, 상기 서브-골 플래너로 하여금, (i) 액션 홀더와, 상기 플레이스 홀더로 오브젝트 홀더 및 리셉터클 홀더를 포함하는 서브-골 프레임을 이용하여, 상기 액션 컴포넌트 시퀀스에 대응되는 제1 서브-골 액션 내지 제N 서브-골 액션 각각을 상기 서브-골 프레임의 상기 액션 홀더에 입력하여 제1 서브-골 프레임 내지 제N 서브-골 프레임을 생성하도록 하고, (ii) 상기 제1 서브-골 액션 내지 상기 제N 서브-골 액션에 따른 타겟 오브젝트와 타겟 리셉터클 사이의 관계 정보를 이용하여 상기 제1 서브-골 프레임 내지 상기 제N 서브-골 프레임 각각에서의 상기 오브젝트 홀더 및 상기 리셉터클 홀더에 대응되는 메타 클래스들을 생성함으로써 상기 서브-골 프레임 시퀀스를 생성하도록 하며, (iii) 상기 컨텍스트들 각각을 상기 서브-골 프레임 시퀀스에 대응되는 각각의 상기 메타 클래스들에 매칭하여 상기 제1 서브-골 내지 상기 제N 서브-골을 생성하도록 하는 것을 특징으로 하는 AI 에이전트가 개시된다.

[0034] 일례로서, 상기 (I) 프로세스에서, 상기 프로세서가, 상기 서브-골 플래너로 하여금 하기 함수에 따라 상기 제1 서브-골 내지 상기 제N 서브-골을 생성하도록 하되,

[0035] $$f_{sub}(l) = \{S_j\}_{j=1}^{N}, \qquad S_j = (A_j, O_j, R_j)$$

[0036] 상기 함수에서, f_{sub}는 상기 서브-골 플래너를 의미하고, l은 상기 자연어 지시 데이터를 의미하며, S_j는 전체 서브-골 중 j번째 서브-골을 의미하고, N은 서브-골의 전체 개수를 의미하며, A_j는 상기 j번째 서브-골에 대응

하는 j번째 서브-골 액션을 의미하고, O_j는 상기 j번째 서브-골에 대응하는 j번째 타겟 오브젝트를 의미하며, R_j는 상기 j번째 서브-골에 대응하는 j번째 타겟 리셉터클을 의미하는 것을 특징으로 하는 AI 에이전트가 개시된다.

[0037] 일례로서, 상기 (I) 프로세스에서, 상기 프로세서가, 상기 서브-골 플래너로 하여금 하기 함수에 따라 상기 컨텍스트를 예측하도록 하되,

[0038] $$f_{ctxt}^{O}(l) = c_O,\ f_{ctxt}^{M}(l) = c_M,\ f_{ctxt}^{R}(l) = c_R$$

[0039] 상기 함수에서, l은 상기 자연어 지시 데이터를 의미하고, f_{ctxt}^{O}는 상기 태스크를 수행함에 있어 메인 타겟으로 결정되는 메인 오브젝트를 의미하며, f_{ctxt}^{M}는 상기 태스크를 수행함에 있어 상기 메인 오브젝트의 컨테이너로 결정되는 컨테이너 오브젝트를 의미하고, f_{ctxt}^{R}는 상기 태스크를 수행함에 있어 상기 메인 오브젝트 및 상기 컨테이너 오브젝트 중 적어도 하나의 목적지 타겟으로 결정되는 목적지 오브젝트를 의미하는 것을 특징으로 하는 AI 에이전트가 개시된다.

[0040] 일례로서, 상기 (I) 프로세스에서, 상기 프로세서가, 상기 서브-골 플래너로 하여금 하기 함수에 따라 상기 서브-골 프레임을 생성하도록 하되,

[0041] $$f_{sf}(l) = \{F_j\}_{j=1}^{N},\ F_j = \left(A_j, \langle O \rangle_j, \langle R \rangle_j\right),\ \langle \cdot \rangle \in E \cup \{x_O, x_M, x_R\}$$

[0042] 상기 함수에서, f_{sf}는 상기 서브-골 프레임을 의미하고, l은 상기 자연어 지시 데이터를 의미하며, F_j는 전체 서브-골 프레임 중 j번째 서브-골 프레임을 의미하고, N은 상기 서브-골 프레임의 전체 개수를 의미하며, A_j는 상기 j번째 서브-골 프레임에 대응하는 j번째 서브-골 액션을 의미하고, $\langle O \rangle_j$는 상기 j번째 서브-골 프레임에 대응하는 j번째 오브젝트 홀더를 의미하며, $\langle R \rangle_j$는 상기 j번째 서브-골 프레임에 대응하는 j번째 리셉터클 홀더를 의미하고, $\langle \cdot \rangle$는 상기 오브젝트 홀더 및 상기 리셉터클 홀더를 포함하는 전체 오브젝트 홀더를 의미하며, E는 상기 전체 환경 내의 전체 오브젝트의 집합을 의미하고, x_O는 상기 태스크를 수행함에 있어 메인 타겟으로 결정되는 메인 오브젝트가 매칭될 제1 메타클래스를 의미하며, x_M은 상기 메인 오브젝트의 컨테이너로 결정되는 컨테이너 오브젝트가 매칭될 제2 메타클래스를 의미하고, x_R은 상기 메인 오브젝트 및 상기 컨테이너 오브젝트 중 적어도 하나의 목적지 타겟으로 결정되는 목적지 오브젝트가 매칭될 제3 메타클래스를 의미하는 것을 특징으로 하는 AI 에이전트가 개시된다.

[0043] 일례로서, 상기 (I) 프로세스에서, 상기 프로세서가, 상기 디테일드 플래너로 하여금 하기 함수에 따라 상기 제1 실행 액션 내지 상기 제T1 실행 액션 및 상기 제1 실행 액션 내지 상기 제Tn 실행 액션을 생성하도록 하되,

[0044] $$f_{dp}^{g}\left((A_j, O_j, R_j)\right) = \{(a_k, o_k)\}_{k=1}^{Tj}$$

[0045] 상기 함수에서, f_{dp}^{g}는 전체 서브-골 액션 중 j번째 서브-골에 대응하는 서브-골 액션 ^{g} - 상기 ^{g}는 상기 j번째 서브-골에 대응하는 서브-골 액션인 A_j와 서로 동일함 - 에 대한 상기 디테일드 플래너를 의미하고, O_j는 상기 j번째 서브-골에 대응하는 j번째 타겟 오브젝트를 의미하며, R_j는 상기 j번째 서브-골에 대응하는 j번째 타겟 리셉터클을 의미하고, a_k는 상기 j번째 서브-골에 대한 k번째 실행 액션을 의미하며, o_k는 상기 j번째 서브-골에 대한 k번째 실행 오브젝트를 의미하고, 상기 T_j는 상기 j번째 서브-골에 대한 전체 실행 액션 개수를 의미하는 것을 특징으로 하는 AI 에이전트가 개시된다.

발명의 효과

[0046] 본 발명은 서브-골 플래너로 하여금 획득된 자연어 지시 데이터를 러닝 연산하도록 하여 자연어 지시 데이터에 대응되는 서브-골 액션들의 액션 컴포넌트 시퀀스 및 액션 컴포넌트 시퀀스에 대응되는 플레이스 홀더 컴포넌트 시퀀스를 포함하는 서브-골 프레임 시퀀스를 생성하도록 하고, 자연어 지시 데이터로부터 획득한 컨텍스트들을 컨텍스트들에 대응되는 서브-골 프레임 시퀀스의 플레이스 홀더에 입력하여 서브-골 프레임 시퀀스에 대응되는 제1 서브-골 내지 제N 서브-골을 생성하도록 하며, 디테일드 플래너로 하여금 제1 서브-골에 대한 제1 실행 액션 내지 제T1 실행 액션, 내지 제N 서브-골에 대한 제1 실행 액션 내지 제Tn 실행 액션을 생성하도록 하는 효과가 있다.

[0047] 또한, 본 발명은 제1 서브-골 내지 제N 서브-골 중 어느 하나인 제j_k 실행 액션에 대응하여, 이미지 퍼셉션 모듈을 통해 제j_k 실행 액션에 따른 특정 오브젝트를 확인하고, 태스크를 수행하는 전체 환경에 대한 시맨틱 스페이셜 맵 상에서의 제j_k 실행 액션의 이전 실행 액션들에 대한 오브젝트들의 상태 정보를 저장하고 있는 환경 인지 메모리 모듈을 참조하여 특정 오브젝트가 유효한 오브젝트인지 확인하며, 특정 오브젝트가 유효한 오브젝트로 확인되면 특정 오브젝트에 대한 제j_k 실행 액션을 수행하며, 제j_k 실행 액션에 따른 특정 오브젝트의 특정 상태 변화 정보를 환경 인지 메모리 모듈에 저장하여 시맨틱 스페이셜 맵을 업데이트하는 효과가 있다.

도면의 간단한 설명

[0048] 본 발명의 실시예의 설명에 이용되기 위하여 첨부된 아래 도면들은 본 발명의 실시예들 중 단지 일부일 뿐이며, 본 발명이 속한 기술분야에서 통상의 지식을 가진 자(이하 "통상의 기술자")에게 있어서는 발명적 작업이 이루어짐 없이 이 도면들에 기초하여 다른 도면들이 얻어질 수 있다.

도 1은 본 발명의 일 실시예에 따른 컨텍스트 인지 플래닝 모듈(CAP) 및 환경 인지 메모리 모듈(EAM)를 포함하는 CAPEAM 모델에 따른 AI 에이전트의 구성을 나타낸 도면이며,

도 2는 본 발명의 일 실시예에 따른 CAP 및 EAM을 포함하는 CAPEAM 모델에 따른 AI 에이전트의 태스크 수행 방법을 나타낸 흐름도이며,

도 3은 본 발명의 일 실시예에 따른 CAP 및 EAM의 세부 구성을 나타낸 도면이며,

도 4는 본 발명의 일 실시예에 따른 서브-골 플래너로부터 자연어 지시 데이터에 대응되는 복수의 서브-골이 생성되는 일련의 과정과 디테일드 플래너로부터 자연어 지시 데이터에 대응되는 복수의 실행 액션이 생성되는 일련의 과정을 나타낸 도면이며,

도 5a는 본 발명의 일 실시예에 따른 타 오브젝트에 의해 외형 중 일부가 차폐된 특정 오브젝트를 유효한 오브젝트로 판단하기 위한 EAM의 제1 모듈인 "Retrospective Object Recognition"가 탑재된 제1 AI 에이전트와 EAM이 탑재되지 않은 제2 AI 에이전트의 차이점을 나타낸 도면이며,

도 5b는 본 발명의 일 실시예에 따른 실행 액션에 따른 특정 오브젝트를 비탐색 타겟 오브젝트로서 결정함으로써 후속 실행 액션에 따른 별도 오브젝트로서 무효한 오브젝트로 판단하기 위한 EAM의 제2 모듈인 "Object Relocation Tracking"이 탑재된 제3 AI 에이전트와 EAM이 탑재되지 않은 제4 AI 에이전트의 차이점을 나타낸 도면이며,

도 5c는 본 발명의 일 실시예에 따른 상태가 변화하는 특정 오브젝트를 유효한 오브젝트로 판단하기 위한 EAM(600)의 제3 모듈인 "Object Location Caching"이 탑재된 제5 AI 에이전트와 EAM(600)이 탑재되지 않은 제6 AI 에이전트의 차이점을 나타낸 도면이다.

발명을 실시하기 위한 구체적인 내용

[0049] 후술하는 본 발명에 대한 상세한 설명은, 본 발명이 실시될 수 있는 특정 실시예를 예시로서 도시하는 첨부 도면을 참조한다. 이들 실시예는 당업자가 본 발명을 실시할 수 있기에 충분하도록 상세히 설명된다. 본 발명의 다양한 실시예는 서로 다르지만 상호 배타적일 필요는 없음이 이해되어야 한다. 예를 들어, 여기에 기재되어 있는 특정 형상, 구조 및 특성은 일 실시예에 관련하여 본 발명의 정신 및 범위를 벗어나지 않으면서 다른 실시예로 구현될 수 있다. 또한, 각각의 개시된 실시예 내의 개별 구성요소의 위치 또는 배치는 본 발명의 정신 및 범위를 벗어나지 않으면서 변경될 수 있음이 이해되어야 한다. 따라서, 후술하는 상세한 설명은 한정적인 의미로서 취하려는 것이 아니며, 본 발명의 범위는, 적절하게 설명된다면, 그 청구항들이 주장하는 것과 균등한 모든

범위와 더불어 첨부된 청구항에 의해서만 한정된다. 도면에서 유사한 참조부호는 여러 측면에 걸쳐서 동일하거나 유사한 기능을 지칭한다.

[0050] 이하, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 본 발명을 용이하게 실시할 수 있도록 하기 위하여, 본 발명의 바람직한 실시예들에 관하여 첨부된 도면을 참조하여 상세히 설명하기로 한다.

[0051] 도 1은 본 발명의 일 실시예에 따른 컨텍스트 인지 플래닝 모듈(500) 및 환경 인지 메모리 모듈(600)를 포함하는 CAPEAM 모델에 따른 AI 에이전트(100)의 구성을 나타낸 도면이다.

[0052] 도 1을 참조하면, AI 에이전트(100)는 컨텍스트 인지 플래닝 모듈(Context-Aware Planning, CAP)(500) 및 환경 인지 메모리 모듈(Environment-Aware Memory, EAM)(600)를 포함할 수 있다. 이 때, CAP(500) 및 EAM(600)의 입출력 및 연산 과정은 각각 통신부(110) 및 프로세서(120)에 의해 이루어질 수 있다. 다만, 도 1에서는 통신부(110) 및 프로세서(120)의 구체적인 연결 관계를 생략하였다. 또한, 메모리(115)는 후술할 여러 가지 인스트럭션들을 저장한 상태일 수 있고, 프로세서(120)는 메모리(115)에 저장된 인스트럭션들을 수행하도록 됨으로써 추후 설명할 프로세스들을 수행하여 본 발명을 수행할 수 있다. 이와 같이 AI 에이전트(100)가 묘사되었다고 하여, AI 에이전트(100)가 본 발명을 실시하기 위한 미디엄, 프로세서 및 메모리가 통합된 형태인 통합 프로세서를 포함하는 경우를 배제하는 것은 아니다. 또한, 메모리(115)와 EAM(600)는 편의상 구분하여 표시하였지만, 경우에 따라 통합된 메모리로 구현될 수도 있을 것이다.

[0053] 다음으로, 본 발명의 일 실시예에 따른 AI 에이전트의 태스크 수행 방법에 대해 구체적으로 설명하도록 한다. 이를 위해 도 2를 참조로 하여 설명하겠다.

[0054] 도 2는 본 발명의 일 실시예에 따른 CAP(500) 및 EAM(600)을 포함하는 CAPEAM 모델에 따른 AI 에이전트(100)의 태스크 수행 방법을 나타낸 흐름도이다.

[0055] 도 2를 참조하면, 자연어 지시 데이터가 획득되면, AI 에이전트(100)가, 상기 자연어 지시 데이터를 서브-골 플래너로 입력하여 서브-골 플래너로 하여금 상기 자연어 지시 데이터를 러닝 연산하도록 하여 상기 자연어 지시 데이터에 대응되는 서브-골 액션들의 액션 컴포넌트 시퀀스 및 상기 액션 컴포넌트 시퀀스에 대응되는 플레이스 홀더 컴포넌트 시퀀스를 포함하는 서브-골 프레임 시퀀스를 생성하도록 하고, 상기 자연어 지시 데이터로부터 획득한 컨텍스트들을 상기 컨텍스트들에 대응되는 상기 서브-골 프레임 시퀀스의 플레이스 홀더에 입력하여 상기 서브-골 프레임 시퀀스에 대응되는 제1 서브-골 내지 제N 서브-골(N은 1 이상의 정수임)을 생성하도록 하며, 디테일드 플래너로 하여금 상기 제1 서브-골에 대한 제1 실행 액션 내지 제T1 실행 액션(T1은 1 이상의 정수임), 내지 상기 제N 서브-골에 대한 제1 실행 액션 내지 제Tn 실행 액션(Tn은 1 이상의 정수임)을 생성하도록 한다(S201).

[0056] 먼저, 서브-골 플래너(510)와 디테일드 플래너(520)에 대한 구체적인 설명을 위해 도 3과 도 4를 참조로 하여 설명한다.

[0057] 도 3은 본 발명의 일 실시예에 따른 CAP(500)및 EAM(600)의 세부 구성을 나타낸 도면이며, 도 4는 본 발명의 일 실시예에 따른 서브-골 플래너(510)로부터 자연어 지시 데이터에 대응되는 복수의 서브-골이 생성되는 일련의 과정과 디테일드 플래너(520)로부터 자연어 지시 데이터에 대응되는 복수의 실행 액션이 생성되는 일련의 과정을 나타낸 도면이다.

[0058] 먼저, 도 3을 참조하면, AI 에이전트(100)는, 자연어 지시 데이터인 "Put the butter knife with the pan in the fridge."가 획득되면, 상기 자연어 지시 데이터를 상기 자연어 지시 데이터에 대응되는 복수의 서브 골을 생성할 서브-골 플래너(510)에 입력한다. 그리고, 서브-골 플래너(510)로 하여금 상기 자연어 지시 데이터를 러닝 연산하도록 하여, 서브-골 플래너(510)의 제1 하위 모듈인 서브-골 프레임 시퀀스 제너레이터(511)로 하여금 상기 자연어 지시 데이터에 대응되는 서브-골 액션들의 액션 컴포넌트 시퀀스 및 상기 서브-골 액션들의 액션 컴포넌트 시퀀스에 대응되는 플레이스 홀더 컴포넌트 시퀀스를 포함하는 서브-골 프레임 시퀀스를 생성하도록 한다.

[0059] 여기서, 서브-골 프레임 시퀀스는, 액션 홀더와, 플레이스 홀더(즉, 오브젝트 홀더 및 리셉터클 홀더)를 포함하는 서브-골 프레임을 이용하여 액션 컴포넌트 시퀀스에 대응되는 제1 서브-골 액션 내지 제N 서브-골 액션 각각을 서브-골 프레임의 액션 홀더에 입력하여 제1 서브-골 프레임 내지 제N 서브-골 프레임을 생성하도록 한다. 좀 더 구체적으로는, AI 에이전트(100)는, 서브-골 프레임 시퀀스 제너레이터(511)로 하여금 하기 함수에 따라 서브-골 프레임을 생성하도록 한다. 참고로, 본 명세서에서 플레이스 홀더는 2 개의 홀더(즉, 오브젝트 홀더 및 리셉터클 홀더)를 포함하는 것으로 예시적으로 설명하였으나, 이에 한정되는 것은 아니며, 3개 이상의 홀더

를 포함하여 구성될 수도 있을 것이다.

[0060] $$f_{sf}(l) = \{F_j\}_{j=1}^{N},\ F_j = \left(A_j, \langle O \rangle_j, \langle R \rangle_j\right),\ \langle \cdot \rangle \in \mathrm{E} \cup \{x_O, x_M, x_R\}$$

[0061] 상기 함수에서, f_{sf}는 상기 서브-골 프레임을 의미하고, l은 상기 자연어 지시 데이터를 의미하며, F_j는 전체 서브-골 프레임 중 j번째 서브-골 프레임을 의미하고, N은 상기 서브-골 프레임의 전체 개수를 의미하며, A_j는 상기 j번째 서브-골 프레임에 대응하는 j번째 서브-골 액션을 의미하고, $\langle O \rangle_j$는 상기 j번째 서브-골 프레임에 대응하는 j번째 오브젝트 홀더를 의미하며, $\langle R \rangle_j$는 상기 j번째 서브-골 프레임에 대응하는 j번째 리셉터클 홀더를 의미하고, $\langle \cdot \rangle$는 상기 오브젝트 홀더 및 상기 리셉터클 홀더를 포함하는 플레이스 홀더를 의미하며, E는 상기 전체 환경 내의 전체 오브젝트의 집합을 의미하고, x_O는 상기 태스크를 수행함에 있어 메인 타겟으로 결정되는 메인 오브젝트(가령, "Butter Knife")가 매칭될 제1 메타클래스를 의미하며, x_M은 상기 메인 오브젝트의 컨테이너로 결정되는 컨테이너 오브젝트(가령, "Pan")가 매칭될 제2 메타클래스를 의미하고, x_R은 상기 메인 오브젝트 및 상기 컨테이너 오브젝트 중 적어도 하나의 목적지 타겟으로 결정되는 목적지 오브젝트(가령, "Fridge")가 매칭될 제3 메타클래스를 의미한다. 여기서, 메인 오브젝트, 컨테이너 오브젝트, 목적지 오브젝트, 제1 메타클래스, 제2 메타클래스, 및 제3 메타클래스에 대한 정확한 설명은 후술하기로 한다. 즉, 서브-골 프레임 시퀀스 제너레이터(511)는, 상기 함수에 따라 상기 자연어 지시 데이터인 "Put the butter knife with the pan in the fridge."에 대응하는 제1 서브-골 프레임인 $F_1 = \left(Pickup_1, \langle O \rangle_1, \langle R \rangle_1\right)$, 제2 서브-골 프레임인 $F_2 = \left(Put_2, \langle O \rangle_2, \langle R \rangle_2\right)$, 제3 서브-골 프레임인 $F_3 = \left(Pickup_3, \langle O \rangle_3, \langle R \rangle_3\right)$, 및 제4 서브-골 프레임인 $F_4 = \left(Put_4, \langle O \rangle_4, \langle R \rangle_4\right)$를 생성할 수 있다.

[0062] 한편, 상기 제1 서브-골 액션 내지 상기 제4 서브-골 액션 각각에 대응되는 오브젝트 홀더 각각 및 리셉터클 홀더 각각은 상기 자연어 지시 데이터로부터 획득한 컨텍스트들이 입력될 수 있는데, 여기서는 좀 더 구체적인 설명을 위해 도 4를 참조로 하여 설명하겠다.

[0063] 도 4를 참조하면, AI 에이전트(100)는, 자연어 지시 데이터가 획득되면, 서브-골 플래너(510)의 제2 하위 모듈인 컨텍스트 프레딕터(512)로 하여금 상기 자연어 지시 데이터에 대응되는 컨텍스트들, 즉, 태스크를 수행함에 있어 메인 타겟으로 결정되는 메인 오브젝트("o: Butter Knife"), 상기 메인 오브젝트의 컨테이너로 결정되는 컨테이너 오브젝트("m: Pan"), 및 상기 메인 오브젝트 및 상기 컨테이너 오브젝트 중 적어도 하나의 목적지 타겟으로 결정되는 목적지 오브젝트("r: Fridge")를 예측하도록 한다. 좀 더 구체적으로는, AI 에이전트(100)는, 컨텍스트 프레딕터(512)로 하여금 하기 함수에 따라 컨텍스트들을 예측하도록 한다.

[0064] $$f_{ctxt}^{O}(l) = c_O,\ f_{ctxt}^{M}(l) = c_M,\ f_{ctxt}^{R}(l) = c_R$$

[0065] 상기 함수에서, l은 상기 자연어 지시 데이터를 의미하고, f_{ctxt}^{O}는 상기 태스크를 수행함에 있어 메인 타겟으로 결정되는 메인 오브젝트를 의미하며, f_{ctxt}^{M}는 상기 태스크를 수행함에 있어 상기 메인 오브젝트의 컨테이너로 결정되는 컨테이너 오브젝트를 의미하고, f_{ctxt}^{R}는 상기 태스크를 수행함에 있어 상기 메인 오브젝트 및 상기 컨테이너 오브젝트 중 적어도 하나의 목적지 타겟으로 결정되는 목적지 오브젝트를 의미한다. 즉, "Put the butter knife with the pan in the fridge."라는 자연어 지시 데이터가 획득되면, AI 에이전트(100)는, 컨텍스트 프레딕터(512)로 하여금, 메인 오브젝트인 c_O로서 "Butter knife"를, 컨테이너 오브젝트인 c_M으로서 "Pan"을, 목적지 오브젝트인 c_R로서 "Fridge"를 예측하도록 할 수 있을 것이다.

[0066] 이와 같이 상기 자연어 지시 데이터에 대한 컨텍스트들이 예측되면, AI에이전트(100)는 상기 자연어 지시 데이터로부터 획득한 컨텍스트들을 상기 컨텍스트들에 대응되는 서브-골 프레임 시퀀스의 플레이스 홀더(즉, 오브젝

트 홀더인 $\langle O \rangle_j$ 및 리셉터클 홀더인 $\langle R \rangle_j$)에 입력하여 서브-골 프레임 시퀀스에 대응되는 제1 서브-골 내지 제N 서브-골을 생성할 수 있다. 한편, AI 에이전트(100)는, 서브-골 플래너(510)로 하여금 특정 함수에 따라 제1 서브-골 내지 제N 서브-골을 생성할 수 있는데 이에 대한 구체적인 설명을 하기 이전에, 먼저 오브젝트 홀더 및 리셉터클 홀더에 대응되는 메타 클래스들에 대해 설명하겠다.

[0067] 도 3과 도 4에는 도시되어 있지는 않지만, 제1 서브-골 프레임 내지 제N 서브-골 프레임 각각에서의 오브젝트 홀더 각각 및 리셉터클 홀더 각각에 생성되는 메타 클래스들은, AI 에이전트(100)가 상기 자연어 지시 데이터에 대응하는 태스크를 수행함에 있어 메인 타겟으로 결정되는 메인 오브젝트가 매칭될 제1 메타 클래스(x_o), 상기 메인 오브젝트의 컨테이너로 결정되는 컨테이너 오브젝트가 매칭될 제2 메타 클래스(x_m), 및 상기 메인 오브젝트의 목적지 타겟 및/또는 상기 컨테이너 오브젝트의 목적지 타겟으로 결정되는 목적지 오브젝트가 매칭될 제3 메타 클래스(x_r)를 포함한다.

[0068] 좀 더 구체적으로는, 서브-골 플래너(510)로부터 제1 서브-골 프레임 내지 제N 서브-골 프레임이 생성되면, AI 에이전트(100)가, 서브-골 플래너(510)로 하여금 제1 서브-골 액션 내지 제N 서브-골 액션에 따른 타겟 오브젝트(즉, 오브젝트 홀더에 입력될 미정의 오브젝트)와 타겟 리셉터클(즉, 리셉터클 홀더에 입력될 미정의 오브젝트) 사이의 관계 정보를 이용하여 제1 서브-골 프레임 내지 제N 서브-골 프레임 각각에서의 오브젝트 홀더 및 리셉터클 홀더에 대응되는 메타 클래스들을 생성함으로써 서브-골 프레임 시퀀스를 생성하도록 할 수 있는데, 여기서, 제1 서브-골 프레임 내지 제N 서브-골 프레임 각각의 오브젝트 홀더에는 제1 메타 클래스 및 제2 메타 클래스 중 적어도 하나가 할당될 수 있고, 상기 제1 서브-골 프레임 내지 상기 제N 서브-골 프레임 각각의 리셉터클 홀더에는 제2 메타 클래스 및 제3 메타 클래스 중 적어도 하나가 할당될 수 있으나, 이에 한정하지는 않는다. 결국, 제1 서브-골 프레임 내지 제N 서브-골 프레임에는, 제1 서브-골 액션 내지 제N 서브-골 액션과 상기 제1 서브-골 프레임 내지 상기 제N 서브-골 프레임 각각의 메타 클래스들 각각이 매칭된 상태일 수 있다.

[0069] 좀 더 구체적인 설명을 위해 다시 도 4를 참조하여 설명하면, 자연어 지시 데이터("Put the butter knife with the pan in the fridge.")에 대응되는 제1 서브-골 프레임 내지 제N 서브-골 프레임 중 특정 서브-골 프레임이 "$F_s = \left(Put, \langle O \rangle_s, \langle R \rangle_s\right)$"로 결정되면, AI 에이전트(100)는, 상기 특정 서브-골 프레임의 특정 서브-골 액션인 "Put"에 따른 타겟 오브젝트 및 타겟 리셉터클 사이의 관계 정보를 이용하여 상기 특정 서브-골 프레임에서의 오브젝트 홀더($\langle O \rangle_s$) 및 리셉터클 홀더($\langle R \rangle_s$) 부분에 메타 클래스(여기서는, 하나의 예시로서, 제2 메타 클래스(x_m) 및 제3 메타 클래스(x_r)를 상정하였지만, 경우에 따라서는 제1 메타 클래스(x_o) 및 제2 메타 클래스(x_m), 또는 제1 메타 클래스(x_o) 및 제3 메타 클래스(x_r)를 상정할 수도 있을 것이다)를 생성함으로써 상기 특정 서브-골 프레임을 "$F_s = (Put, x_m, x_r)$" 상태로 만들 것이다. 그리고, 서브-골 플래너(510)는 상기 자연어 지시 데이터로부터 컨텍스트 프레딕터(512)에 의해 획득된 컨텍스트들("o: Butter Knife", "m: Pan", "r: Fridge") 중, 제2 메타 클래스에 대응되는 "m: Pan"을 상기 특정 서브-골 프레임의 "x_m" 부분에 매칭하고, 제3 메타 클래스에 대응되는 "r: Fridge"를 상기 특정 서브-골 프레임의 "x_r" 부분에 매칭함으로써 특정 서브-골"($S_s = (Put, Pan, Fridge)$" 을 생성할 수 있는데, 서브-골을 생성하는 좀 더 구체적인 방법은 후술하기로 한다.

[0070] 한편, 리셉터클 홀더에는 메타 클래스가 생성되지 않는 경우가 존재하는데, 이에 대한 좀 더 구체적인 설명을 위해 도 3을 다시 참조하여 설명하자면, AI 에이전트(100)는, 제1 서브-골 액션인 "Pickup"을 포함하는 제1 서브-골 프레임($F_1 = \left(Pickup_1, \langle O \rangle_1, \langle R \rangle_1\right)$)의 제1 서브-골 액션인 "Pickup"에 따른 타겟 오브젝트(가령, "Butter Knife")가 입력되기 이전에, 오브젝트 홀더에 제1 메타 클래스(노랑 영역에 해당함)를 생성하고, 리셉터클 홀더에는 메타 클래스를 생성하지 않는다는 의미("-" 표시되어 있으며 NULL 의 의미)로서, 리셉터클 홀더에는 메타 클래스가 생성되지 않은 상태인 제1 서브-골 프레임($F_1 = \left(Pickup_1, \langle x_o \rangle_1, \langle - \rangle_1\right)$)을 완성할 수 있을 것이다. 반면에, AI 에이전트(100)는, 제2 서브-골 액션인 "Put"를 포함하는 제2 서브-골 프레임

($F_2 = \left(Put_2, \langle O \rangle_2, \langle R \rangle_2\right)$)의 제2 서브-골 액션인 "Put"에 따른 타겟 오브젝트(가령, "Butter Knife") 및 타겟 리셉터클(가령, "Pan")이 입력되기 이전에, 오브젝트 홀더에 제1 메타 클래스(노랑 영역에 해당함)를 생성하고, 리셉터클 홀더에는 제2 메타 클래스(파랑 영역에 해당함)를 생성함으로써, 제2 서브-골 프레임 ($F_2 = \left(Put_2, \langle x_O \rangle_2, \langle x_m \rangle_2\right)$)을 완성할 수 있을 것이다. 이와 같은 방법으로, AI 에이전트(100)는, 제3 서브-골 프레임 및 제4 서브-골 프레임 각각 역시 ($F_3 = \left(Pickup_3, \langle x_m \rangle_3, \langle - \rangle_3\right)$) 및 ($F_4 = \left(Put_4, \langle x_m \rangle_4, \langle x_r \rangle_4\right)$)로서 완성할 수 있을 것이다.

[0071] 다음으로, AI 에이전트(100)는, 상기 자연어 지시 데이터로부터 획득한 컨텍스트들 각각을 서브-골 프레임 시퀀스에 대응되는 각각의 메타 클래스들에 매칭하여 제1 서브-골 내지 제N 서브-골을 생성한다. 좀 더 구체적으로는, AI 에이전트(100)는, 서브-골 플래너(510)로 하여금 하기 함수에 따라 제1 서브-골 내지 제N 서브-골을 생성하도록 한다.

[0072] $$f_{sub}(l) = \{S_j\}_{j=1}^{N}, \qquad S_j = (A_j, O_j, R_j)$$

[0073] 상기 함수에서, f_{sub}는 상기 서브-골 플래너(510)를 의미하고, l은 상기 자연어 지시 데이터를 의미하며, S_j는 전체 서브-골 중 j번째 서브-골을 의미하고, N은 서브-골의 전체 개수를 의미하며, A_j는 상기 j번째 서브-골에 대응하는 j번째 서브-골 액션을 의미하고, O_j는 상기 j번째 서브-골에 대응하는 j번째 타겟 오브젝트를 의미하며, R_j는 상기 j번째 서브-골에 대응하는 j번째 타겟 리셉터클을 의미한다. 즉, 서브-골 플래너(510)는, 상기 제1 서브-골 프레임($F_1 = \left(Pickup_1, \langle x_O \rangle_1, \langle - \rangle_1\right)$)에 대응되는 메타 클래스("$x_O$")에 상기 자연어 지시 데이터로부터 획득한 컨텍스트(가령, "Butter Knife")를 매칭하여 제1 서브-골인 $S_1 = (Pickup_1, Butter\ Knife_1, -)$을 생성하고, 상기 제2 서브-골 프레임($F_2 = \left(Put_2, \langle x_O \rangle_2, \langle x_m \rangle_2\right)$)에 대응되는 메타 클래스들("$x_O$", "$x_m$")에 상기 자연어 지시 데이터로부터 획득한 컨텍스트들(가령, "Butter Knife", "Pan")을 매칭하여 제2 서브-골인 $S_2 = (Put_2, Butter\ Knife_2, Pan_2)$를 생성하며, 상기 제3 서브-골 프레임($F_3 = \left(Pickup_3, \langle x_m \rangle_3, \langle - \rangle_3\right)$)에 대응되는 메타 클래스("$x_m$")에 상기 자연어 지시 데이터로부터 획득한 컨텍스트(가령, "Pan")를 매칭하여 제3 서브-골인 $S_3 = (Pickup_3, Pan_3, -)$를 생성하고, 상기 제4 서브-골 프레임($F_4 = \left(Put_4, \langle x_m \rangle_4, \langle x_r \rangle_4\right)$)에 대응되는 메타 클래스들("$x_m$", "$x_r$")에 상기 자연어 지시 데이터로부터 획득한 컨텍스트들(가령, "Pan", "Fridge")을 매칭하여 제4 서브-골인 $S_4 = (Put_4, Pan_4, Fridge_4)$를 생성할 수 있다.

[0074] 다음으로, 서브-골 플래너(510)로부터 생성된 복수의 서브-골 각각에 대한 구체적인 실행 액션을 생성하는 디테일드 플래너(520)에 대하여 설명하겠다.

[0075] 먼저, AI 에이전트(100)는, 디테일드 플래너(520)로 하여금 제1 서브-골에 대한 제1 실행 액션 내지 제T1 실행 액션(여기서, T1은 1 이상의 정수임), 내지 제N 서브-골에 대한 제1 실행 액션 내지 제Tn 실행 액션(여기서, Tn은 1 이상의 정수임)을 생성하도록 할 수 있다. 좀 더 구체적으로는, AI 에이전트(100)는, 디테일드 플래너(520)로 하여금 하기 함수에 따라 제1 실행 액션 내지 제T1 실행 액션을 생성하도록 하는 프로세스 내지 제1 실행 액션 내지 제Tn 실행 액션을 생성하도록 하는 프로세스를 수행하도록 할 수 있다. 한편, T1과 Tn은 서로 같거나 다른 정수를 갖을 수 있을 것이다.

[0076] $$f_{dp}^{g}\left((A_j, O_j, R_j)\right) = \{(a_k, o_k)\}_{k=1}^{Tj}$$

[0077] 상기 함수에서, f_{dp}^{g}는 전체 서브-골 액션 중 j번째 서브-골에 대응하는 서브-골 액션 g (상기 g 는 상기 j번째 서브-골에 대응하는 서브-골 액션인 A_j와 서로 동일함)에 대한 상기 디테일드 플래너(520)를 의미하고, O_j는 상기 j번째 서브-골에 대응하는 j번째 타겟 오브젝트를 의미하며, R_j는 상기 j번째 서브-골에 대응하는 j번째 타겟 리셉터클을 의미하고, a_k는 상기 j번째 서브-골에 대한 k번째 실행 액션을 의미하며, o_k는 상기 j번째 서브-골에 대한 k번째 실행 오브젝트를 의미하고, 상기 T_j는 상기 j번째 서브-골에 대한 전체 실행 액션 개수를 의미한다.

[0078] 좀 더 구체적인 설명을 위해 도 3을 다시 참조하면, 디테일드 플래너(520)는, 상기 함수에 따라 상기 자연어 지시 데이터인 "Put the butter knife with the pan in the fridge."에 대응하는 제1 서브-골 내지 제4 서브-골 각각에 대한 실행 액션들을 생성할 수 있는데, 일례로, 제1 서브-골($S_1 = (Pickup_1, Butter\ Knife_1, -)$)에 대해서는, 제1 실행 액션 $\{(Find_1, Butter\ Knife_1)\}^{T1}$(미도시) 및 제2 실행 액션 $\{(Pickup_2, Butter\ Knife_2)\}^{T1}$(미도시)를 생성할 수 있고, 제2 서브-골($S_2 = (Put_2, Butter\ Knife_2, Pan_2)$)에 대해서는, 제1 실행 액션 $\{(Find_1, Pan_1)\}^{T2}$(미도시) 및 제2 실행 액션$\{(Put_2, Pan_2)\}^{T2}$(미도시)를 생성할 수 있으며, 제3 서브-골($S_3 = (Pickup_3, Pan_3, -)$)에 대해서는, 제1 실행 액션 $\{(Pickup_1, Pan_1)\}^{T3}$(미도시)을 생성할 수 있고, 제4 서브-골($S_4 = (Put_4, Pan_4, Fridge_4)$)에 대해서는, 제1 실행 액션 $\{(Open_1, Fridge_1)\}^{T4}$, 제2 실행 액션 $\{(Put_2, Pan_2)\}^{T4}$, 및 제3 실행 액션 $\{(Close_3, Fridge_3)\}^{T4}$을 생성할 수 있다. 만약, 상기 제4 서브-골에 대한 제1 실행 액션을 AI 에이전트(100)가 수행하게 된다면, AI 에이전트(100)는 상기 제3 서브-골에 대한 제1 실행 액션에 따른 "Pan"을 "Pickup"한 상태에서, "Fridge"의 문을 "Open"하는 상기 제4 서브-골에 대한 제1 실행 액션을 수행할 것이다. 또한, 또 다른 예로서, 도 4를 다시 참조하여 설명하면, 특정 서브-골($S_s = (Put, Pan, Fridge)$이 결정된 상태에서 상기 특정 서브-골에 대한 특정 실행 액션을 AI 에이전트(100)가 수행하게 된다면, AI 에이전트(100)는 "Pan"을 "Pickup"한 상태에서, "Fridge"의 문을 "Open" 한 뒤(미도시), 상기 "Pan"을 "Fridge" 내부에 "Put"하는 특정 실행 액션$\{(Put, Fridge)\}^{Ts}$, 을 수행할 수 있을 것이다.

[0079] 다음으로, EAM(600)을 가지는 AI에이전트(100)가, 자연어 지시 데이터에 대응하는 제1 실행 액션 내지 제T 실행 액션 각각을 수행하는 일련의 과정에 대하여 설명하도록 한다.

[0080] 다시 도 2 를 참조하면, S201단계에 후속하여, AI 에이전트(100)는, 제1 서브-골 내지 제N 서브-골 중 어느 하나인 제j 서브-골에 대한 제k 실행 액션을 제j_k 실행 액션이라고 할 때, (i) 상기 j를 1부터 상기 N까지 증가시키고, 상기 k를 1부터 상기 Tj까지 증가시켜가며, 상기 제j_k 실행 액션에 대응하여, 이미지 퍼셉션 모듈을 통해 상기 제j_k 실행 액션에 따른 특정 오브젝트를 확인하고, (ii) 태스크를 수행하는 전체 환경에 대한 시맨틱 스페이셜 맵 상에서의 상기 제j_k 실행 액션의 이전 실행 액션들에 대한 오브젝트들의 상태 정보를 저장하고 있는 환경 인지 메모리 모듈(600)를 참조하여 상기 특정 오브젝트가 유효한 오브젝트인지 확인하며, 상기 특정 오브젝트가 유효한 오브젝트로 확인되면 상기 특정 오브젝트에 대한 상기 제j_k 실행 액션을 수행하며, 상기 제j_k 실행 액션에 따른 상기 특정 오브젝트의 특정 상태 변화 정보를 상기 환경 인지 메모리 모듈(600)에 저장하여 상기 시맨틱 스페이셜 맵을 업데이트한다(S202).

[0081] 여기서, 시맨틱 스페이셜 맵은, AI 에이전트(100)가, 이미지 퍼셉션 모듈로부터 획득된 전체 환경에 대한 공간 정보를 참조로 하여 전체 환경에 대응하는 뎁스-맵(depth map)을 생성하고, 이미지 퍼셉션 모듈로부터 전체 환경 내의 전체 오브젝트 중 적어도 일부 각각에 대응하는 전체 오브젝트 마스크 각각을 획득하며, 전체 오브젝트 마스크 각각과 뎁스-맵을 3D 세계 좌표로 백프로젝팅함으로써 구축된다.

[0082] 또한, AI 에이전트(100)는 EAM(600)에 저장된 특정 오브젝트에 대한 특정 상태 변화 정보(여기서, 특정 상태 변화 정보는, 상기 제j_k 실행 액션을 수행하기 이전에 획득된 상기 특정 오브젝트의 특정 최초 위치 정보, 상기 제j_k 실행 액션을 수행하기 이후에 획득된 상기 특정 오브젝트의 특정 최신 위치 정보, 상기 제j_k 실행 액션을 수행하기 이전에 획득된 상기 특정 오브젝트의 특정 최초 마스크 정보, 및 상기 제j_k 실행 액션을 수행하기 이후에 획득된 상기 특정 오브젝트의 특정 최신 마스크 정보 중 적어도 일부를 참조로 하여 결정됨)를 참조로 하여 시맨틱 스페이셜 맵을 실시간으로 업데이트 할 수 있다. 다음으로, EAM(600)의 제1 모듈인 "Retrospective Object Recognition"(610), 제2 모듈인 "Object Relocation Tracking"(620), 및 제3 모듈인 "Object Location

Caching"(630)에 대한 특징을 설명하겠다.

[0083] 먼저, AI 에이전트(100)는, 상기 특정 오브젝트의 상기 특정 최신 마스크 정보를 상기 이미지 퍼셉션 모듈로부터 획득하고 상기 특정 최신 마스크 정보를 상기 환경 인지 메모리 모듈(600)에 저장한 상태에서, 상기 제j_k 실행 액션의 제1 후속 실행 액션에 따른 타 오브젝트에 의해 상기 특정 오브젝트의 외형 중 일부가 차폐됨으로써 제2 후속 실행 액션에 따른 상기 특정 오브젝트의 특정 차폐 마스크 정보(여기서, 특정 차폐 마스크 정보는 상기 이미지 퍼셉션 모듈로부터 획득된다)와 상기 특정 최신 마스크 정보가 서로 불일치하는 것으로 판단되면, 환경 인지 메모리 모듈(600)를 참조로 하여, 상기 특정 오브젝트의 상기 특정 최신 위치 정보를 파악한 다음, 상기 제2 후속 실행 액션에 따른 상기 특정 오브젝트의 상기 특정 차폐 마스크 정보를 상기 특정 최신 마스크 정보로 대체함으로써 상기 특정 오브젝트를 유효한 오브젝트로 판단하고, 상기 특정 오브젝트에 대한 상기 제2 후속 실행 액션을 수행할 수 있다. 여기서는 좀 더 구체적인 설명을 위해 도 5a를 참조로 하여 설명한다.

[0084] 도 5a는 본 발명의 일 실시예에 따른 타 오브젝트에 의해 외형 중 일부가 차폐된 특정 오브젝트를 유효한 오브젝트로 판단하기 위한 EAM(600)의 제1 모듈인 "Retrospective Object Recognition"(610)이 탑재된 제1 AI 에이전트와 EAM(600)이 탑재되지 않은 제2 AI 에이전트의 차이점을 나타낸 도면이다.

[0085] 도 5a를 참조하면, 자연어 지시 데이터인 "Put the watch from shelf in the blue bowl move the whole bowl to the tv cabinet."이 획득되면, EAM(600)의 제1 모듈을 탑재한 제1 AI 에이전트와, EAM(600)을 탑재하지 않은 제2 AI 에이전트는, CAP(500)로 하여금 상기 자연어 지시 데이터에 대응되는 복수의 서브-골과 상기 복수의 서브-골 각각에 대한 적어도 하나의 실행 액션을 생성하도록 할 수 있으나, 태스크를 수행하는 과정에서 동일한 특정 오브젝트에 대한 유효성은 상황에 따라 서로 다르게 판단하는 것을 확인할 수 있다.

[0086] 좀 더 구체적으로는, 제1 AI 에이전트 및 제2 AI 에이전트 모두가 상기 자연어 지시 데이터에 대응하는 제1_1 실행 액션을 ("Put", "Bowl")로서 생성하고, 제2_1 실행 액션을 ("Pick", "Bowl")로서 생성하였다고 할 때, 제2 AI 에이전트는, "Watch"를 "Bowl"에 "Put"하는 제1_1 실행 액션에 대해서는 성공적으로 수행하였지만, "Watch"에 의해 외형 중 일부가 차폐된 상태의 "Bowl"을 "Pick"하는 제2_1 실행 액션에 대해서는 성공적으로 수행하지 못하는 것을 확인할 수 있다.

[0087] 반면에, 제1 AI 에이전트는, "Watch"를 "Bowl"에 "Put"하는 제1_1 실행 액션과 "Watch"에 의해 외형 중 일부가 차폐된 상태의 "Bowl"을 "Pick"하는 제2_1 실행 액션을 모두 성공적으로 수행하였음을 확인할 수 있는데, 이는, 제1 모듈을 탑재한 제1 AI 에이전트가, 제1_1 실행 액션을 수행하기 이전의 "Bowl"의 최신 마스크 정보를 이미지 퍼셉션 모듈로부터 획득하여 EAM(600)에 저장하고, 제1_1 실행 액션을 수행한 후의 "Bowl"의 차폐 마스크 정보와 "Bowl"의 최신 마스크 정보를 비교하여 서로 불일치하는 것으로 판단되면, 제2_1 실행 액션에 따른 "Bowl"의 차폐 마스크 정보를 "Bowl"의 최신 마스크 정보로 대체함으로써 "Bowl"을 유효한 오브젝트로 판단하며, 이에 따라 제1 AI 에이전트는 "Watch"에 의해 "Bowl"의 외형 중 일부가 차폐되더라도 "Bowl"에 대한 제2_1 실행 액션을 성공적으로 수행하기 때문이다. 다만, 이에 한정되는 것은 아니며, 제1_1 실행 액션을 수행한 후에 곧바로 "Bowl"의 차폐 마스크 정보를 획득하여 이를 EAM(600)에 저장할 수도 있을 것이다.

[0088] 한편, 위에서는 제1_1 실행 액션을 ("Put", "Bowl")로서 생성하고, 제2_1 실행 액션을 ("Pick", "Bowl")로서 생성하는 것을 가정하였지만, 이에 한정하지 않으며, 경우에 따라서는 제1_1 실행 액션을 ("Put", "Bowl")로서 생성하고, 제1_2 실행 액션을 ("Pick", "Bowl")로서 생성할 수도 있을 것이다.

[0089] 다음으로, AI 에이전트(100)는, 상기 특정 오브젝트와의 특정 인터랙션과, 상기 특정 오브젝트와 동일한 클래스를 가지는 적어도 하나의 별도 오브젝트와의 별도 인터랙션이 상기 자연어 지시 데이터에 포함되는 것으로 판단되면, 상기 제j_k 실행 액션의 제3 후속 실행 액션에 따른 상기 별도 인터랙션을 수행하기 이전에, 상기 제j_k 실행 액션에 따른 상기 특정 오브젝트를 비탐색 타겟 오브젝트로서 결정하고, 상기 이미지 퍼셉션 모듈로부터 획득된 상기 특정 최신 위치 정보를 참조로 하여 상기 특정 오브젝트를 상기 제3 후속 실행 액션에 따른 상기 별도 오브젝트로서 무효한 오브젝트로 결정할 수 있다. 여기서는 좀 더 구체적인 설명을 위해 도 5b를 참조로 하여 설명한다.

[0090] 도 5b는 본 발명의 일 실시예에 따른 실행 액션에 따른 특정 오브젝트를 비탐색 타겟 오브젝트로서 결정함으로써 후속 실행 액션에 따른 별도 오브젝트로서 무효한 오브젝트로 판단하기 위한 EAM(600)의 제2 모듈인 "Object Relocation Tracking"이 탑재된 제3 AI 에이전트와 EAM(600)이 탑재되지 않은 제4 AI 에이전트의 차이점을 나타낸 도면이다.

[0091] 도 5b를 참조하면, 자연어 지시 데이터인 "Place the two TissueBoxes from the shelf on the toilet basin."

이 획득되면, EAM(600)의 제2 모듈을 탑재한 제3 AI 에이전트와, EAM(600)을 탑재하지 않은 제4 AI 에이전트는, CAP(500)로 하여금 상기 자연어 지시 데이터에 대응되는 복수의 서브-골과 상기 복수의 서브-골 각각에 대한 적어도 하나의 실행 액션을 생성하도록 할 수 있으나, 태스크를 수행하는 과정에서 동일 클래스를 갖는 복수의 오브젝트 각각에 대한 유효성은 상황에 따라 서로 다르게 판단하는 것을 확인할 수 있다.

[0092] 좀 더 구체적으로는, 제3 AI 에이전트 및 제4 AI 에이전트 모두가 상기 자연어 지시 데이터에 대응하는 제1_1 실행 액션을 ("Pick", "TissueBox")로서 생성하고, 제2_1 실행 액션을 ("Put", "TissueBox")로서 생성하며, 제3_1 실행 액션을 ("Pick", "TissueBox")로서 생성하였다고 할 때, 제4 AI 에이전트는, 선반 위에 위치하는 두 개의 "TissueBox" 중 하나인 제1 티슈 박스(가령, 특정 오브젝트)를 "Pick"하는 제1_1 실행 액션 및 상기 제1 티슈 박스를 변기 위에 "Put" 하는 제2_1 실행 액션에 대해서는 성공적으로 수행하였지만, 선반 위에 위치하는 나머지 하나의 제2 티슈 박스(가령, 별도 오브젝트)를 "Pick"하는 제3_1 실행 액션에 대해서는 이미 변기 위에 위치하는 제2_1 실행 액션에 따른 제1 티슈 박스를 다시 "Pick"함으로써 태스크를 성공적으로 수행하지 못하는 것을 확인할 수 있다.

[0093] 반면에, 제3 AI 에이전트는, 제1_1 실행 액션, 제2_1 실행 액션, 및 제3_1 실행 액션 모두를 성공적으로 수행하였음을 확인할 수 있는데, 이는, EAM(600)을 탑재한 제3 AI 에이전트가, 제3_1 실행 액션을 수행하기 이전에, 제2_1 실행 액션에 따른 이미 변기 위에 위치하는 제1 티슈 박스의 최신 위치 정보를 EAM(600)에 기록함과 동시에 제1 티슈 박스를 비탐색 타겟 오브젝트로서 결정하고, 제1 티슈 박스에 대한 최신 위치 정보를 참조로 하여 제1 티슈 박스를 제3_1 실행 액션에 따른 제2 티슈 박스로서 무효한 오브젝트로 결정함으로써, 제3_1 실행 액션을 성공적으로 수행하기 때문이다.

[0094] 한편, 위에서는 제1_1 실행 액션을 ("Pick", "TissueBox")로서 생성하고, 제2_1 실행 액션을 ("Put", "TissueBox")로서 생성하며, 제3_1 실행 액션을 ("Pick", "TissueBox")로서 생성하는 것을 가정하였지만, 이에 한정하지 않으며, 경우에 따라서는 제1_1 실행 액션을 ("Pick", "TissueBox")로서 생성하고, 제1_2 실행 액션을 ("Put", "TissueBox")로서 생성하며, 제2_1 실행 액션을 ("Pick", "TissueBox")로서 생성할 수도 있을 것이고, 또 다른 경우에는 제1_1 실행 액션을 ("Pick", "TissueBox")로서 생성하고, 제2_1 실행 액션을 ("Put", "TissueBox")로서 생성하며, 제2_2 실행 액션을 ("Pick", "TissueBox")로서 생성할 수도 있을 것이며, 또 다른 경우에는 제1_1 실행 액션을 ("Pick", "TissueBox")로서 생성하고, 제1_2 실행 액션을 ("Put", "TissueBox")로서 생성하며, 제1_3 실행 액션을 ("Pick", "TissueBox")로서 생성할 수도 있을 것이다.

[0095] 다음으로, AI 에이전트(100)는, 상기 이미지 퍼셉션 모듈로부터 상기 특정 오브젝트의 상기 특정 최신 마스크 정보를 획득하고, 상기 특정 최신 마스크 정보 및 상기 특정 오브젝트의 특정 최신 위치 정보를 상기 환경 인지 메모리 모듈에 저장하며, 상기 제j_k 실행 액션의 제4 후속 실행 액션에 따른 상기 특정 오브젝트에 대한 추가 인터랙션이 결정되면, 상기 특정 최신 위치 정보를 참조로 하여 상기 특정 오브젝트에 대한 최적의 탐색 경로를 결정하고, 상기 특정 최신 마스크 정보를 참조로 하여 상기 특정 오브젝트가 상기 제4 후속 실행 액션에 대해 유효한 오브젝트인지 확인할 수 있다. 여기서는 좀 더 구체적인 설명을 위해 도 5c를 참조로 하여 설명한다.

[0096] 도 5c는 본 발명의 일 실시예에 따른 상태가 변화하는 특정 오브젝트를 유효한 오브젝트로 판단하기 위한 EAM(600)의 제3 모듈인 "Object Location Caching"이 탑재된 제5 AI 에이전트와 EAM(600)이 탑재되지 않은 제6 AI 에이전트의 차이점을 나타낸 도면이다.

[0097] 도 5c를 참조하면, 자연어 지시 데이터인 "Cut two apples into sections move one section from the sink to the counter."가 획득되면, EAM(600)의 제3 모듈을 탑재한 제5 AI 에이전트와, EAM(600)을 탑재하지 않은 제6 AI 에이전트는, CAP(500)로 하여금 상기 자연어 지시 데이터에 대응되는 복수의 서브-골과 상기 복수의 서브-골 각각에 대한 적어도 하나의 실행 액션을 생성하도록 할 수 있으나, 태스크를 수행하는 과정에서 상태가 변화된 특정 오브젝트에 대한 유효성은 상황에 다라 서로 다르게 판단하는 것을 확인할 수 있다.

[0098] 좀 더 구체적으로는, 제5 AI 에이전트 및 제6 AI 에이전트 모두가 상기 자연어 지시 데이터에 대응하는 제1_1 실행 액션을 ("Slice", "Apple")로서 생성하고, 제2_1 실행 액션을 ("Put", "CounterTop")로서 생성하며, 제3_1 실행 액션을 ("Pick", "AppleSliced")로서 생성하고, 제4_1 실행 액션을 ("Put", "CounterTop")로서 생성하며, 제5_1 실행 액션을 ("Pick", "AppleSliced")로서 생성하였다고 할 때, 제6 AI 에이전트는, "Apple"을 "Slice"하는 제1_1 실행 액션과, 슬라이스 된 사과의 일부분을 "CounterTop"에 "Put"하는 제2_1 실행 액션에 대해서는 성공적으로 수행하였지만, 슬라이스 된 사과의 다른 부분을 "Pick"하는 제3_1 실행 액션에 대해서는 성공적으로 수행하지 못하는 것을 확인할 수 있다.

[0099] 반면에, 제5 AI 에이전트는, 제1_1 실행 액션, 제2_1 실행 액션, 제3_1 실행 액션, 제4_1 실행 액션 및 제5_1 실행 액션 모두를 성공적으로 수행하였음을 확인할 수 있는데, 이는, EAM(600)을 탑재한 제5 AI 에이전트가, 제1_1 실행 액션을 수행한 상태에서, 슬라이스 된 사과의 최신 마스크 정보와 최신 위치 정보를 EAM(600)에 저장하고, 제3_1 실행 액션을 수행하기 이전에 상기 최신 위치 정보를 참조로 하여 상기 슬라이스 된 사과에 대한 인터랙션을 하기 위한 최적의 탐색 경로를 결정하며, 상기 최신 마스크 정보를 참조로 하여 상기 슬라이스 된 사과가 유효한 오브젝트인지 확인하고, 상기 슬라이스 된 사과가 유효한 오브젝트로 판단되면, 제3_1 실행 액션을 성공적으로 수행하기 때문이다.

[0100] 한편, 위에서는 제1_1 실행 액션을 ("Slice", "Apple")로서 생성하고, 제2_1 실행 액션을 ("Put", "CounterTop")로서 생성하며, 제3_1 실행 액션을 ("Pick", "AppleSliced")로서 생성하고, 제4_1 실행 액션을 ("Put", "CounterTop")로서 생성하며, 제5_1 실행 액션을 ("Pick", "AppleSliced")로서 생성하는 것을 가정하였지만, 이에 한정하지 않으며, 경우에 따라서는 제1_1 실행 액션을 ("Slice", "Apple")로서 생성하고, 제1_2 실행 액션을 ("Put", "CounterTop")로서 생성하며, 제1_3 실행 액션을 ("Pick", "AppleSliced")로서 생성하고, 제1_4 실행 액션을 ("Put", "CounterTop")로서 생성하며, 제1_5 실행 액션을 ("Pick", "AppleSliced")로서 생성하는 등, 상기 제1_1 실행 액션, 상기 제2_1 실행 액션, 상기 제3_1 실행 액션, 상기 제4_1 실행 액션, 및 상기 제5_1 실행 액션 각각의 서브-골 넘버링 및 실행 액션 넘버링은 다르게 설정될 수 있을 것이다.

[0101] 다음으로, 이하 본 발명의 일 실시예에 따른 컨텍스트 인지 플래닝 모듈(500) 및 환경 인지 메모리 모듈(600)를 포함하는 CAPEAM 모델에 따른 AI 에이전트(100)의 우수성에 대해 설명하도록 한다.

Model	Low Inst.	Tem. Act.	Test Seen SR	Test Seen GC	Test Unseen SR	Test Unseen GC
FILM [27]	✗	✓	25.77 (10.39)	36.15 (14.17)	24.46 (9.67)	34.75 (13.13)
Prompter [19]	✗	✓	**49.38 (23.47)**	**55.90 (29.06)**	42.64 (19.49)	**59.55 (25.00)**
CAPEAM*	✗	✓	46.64 (20.81)	55.29 (25.47)	**45.72 (20.15)**	57.25 (24.73)
FILM [27]	✓	✓	28.83 (11.27)	39.55 (15.59)	27.80 (11.32)	38.52 (15.13)
Prompter [19]	✓	✓	**53.23 (25.81)**	**63.43 (30.72)**	45.72 (20.76)	58.76 **(26.22)**
CAPEAM*	✓	✓	50.62 (22.61)	59.40 (27.49)	**49.84 (22.61)**	**61.10 (27.00)**
HLSM [5]	✗	✗	25.11 (6.69)	35.79 (11.53)	16.29 (4.34)	27.24 (8.45)
LGS-RPA [29]	✗	✗	33.01 (16.65)	41.71 **(24.49)**	27.80 (12.92)	38.55 (20.01)
EPA [26]	✗	✗	39.96 (2.56)	44.14 (3.47)	36.07 (2.92)	39.54 (3.91)
CAPEAM (Ours)	✗	✗	**47.36 (19.03)**	**54.38** (23.78)	**43.69 (17.64)**	**54.66 (22.76)**
HLSM [5]	✓	✗	29.94 (8.74)	41.21 (14.58)	20.27 (5.55)	30.31 (9.99)
MAT [20]	✓	✗	33.01 (9.42)	43.65 (14.68)	21.84 (6.13)	32.41 (10.59)
AMSLAM [21]	✓	✗	29.48 (3.28)	40.88 (5.56)	23.48 (2.36)	34.64 (4.63)
LGS-RPA [29]	✓	✗	40.05 (21.28)	48.66 **(28.97)**	35.41 **(22.76)**	45.24 (22.76)
CAPEAM (Ours)	✓	✗	**51.79 (21.60)**	**60.50** (25.88)	**46.11** (19.45)	**57.33 (24.06)**
Human			-	-	91.00 (85.80)	94.50 (87.60)

[0102]

[0103] 위 실험 결과는 가상 공간 상에서 동작하는 ALFRED 벤치마크를 사용한 것으로, Test Seen은 test시의 데이터셋이 training 시의 데이터셋의 부분집합인 경우이고, Test Unseen은 양 시점의 데이터셋이 다른 경우를 의미한다. 또한, Low Inst. 부분은 서브-골 인스트럭션의 적용 여부를 나타내고, Tem Act. 부분은, So et al. 의 Film: Following instructions in laguage with modular methods(2022)에서 설계된 템플릿된 액션 시퀀스의 적용 여부를 나타내며, SR(success rate) 부분은 지시에 따라 AI 에이전트가 수행하도록 의도된 최종 목표의 완료율을 의미하고, GC(goal-condition)는 각 태스크에 대해서 지시를 구성하는 각각의 작업들(가령, 제1 서브-골 내지 제N 서브-골 각각)의 완료율을 의미하며, CAPEAM*은 상기 템플릿 액션이 적용되지 않은 CAPEAM 모델을 의미하고, 볼드 부분은 매트릭스 각각의 가장 높은 SR 값 및 가장 높은 GC 값을 의미하며, 밑줄 부분은 매트릭스 각각의 두번째로 높은 SR 값 및 두번째로 높은 GC 값을 의미하고, 각 수치에 대응하는 괄호 부분은 AI 에이전트가 수행한 액션의 길이에 따른 SR 및 GC에 패널티를 부여하는 path-length-weighted(PLW) 스코어를 의미한다.

[0104] 위 표에서 볼 수 있듯, Test Seen 부분과 Test Unseen 부분에서 본 발명의 CAPEAM 모델을 적용한 것이 높은 SR 값과 높은 GR 값을 기록했음을 알 수 있다. 한편, Prompter 모델의 경우, Test Seen 부분에서 본 발명의 CAPEAM*을 적용한 것보다 높은 SR 값(49.38; 53.23)과 GR 값(55.90; 63.43)을 기록한 것을 확인할 수 있는데,

이는, Prompter 모델이 Seen Envrionment에 대한 학습된 데이터에 대해서는 성능이 뛰어나지만, 학습에 사용되지 않은 Unseen Envrionment에 대해서는 오버피팅이 발생되었음을 의미한다. 반면에, 본 발명의 CAPEAM 모델은 Test Seen과 Test Unseen 사이에서의 SR 값(46.64, 43.72; 50.62, 49.84)과 GR 값(55.29, 57.25; 59.40, 61.10)의 차이가 크지 않은 것을 확인할 수 있다. 또한, Prompter 모델이 Test Unseen 부분에서 본 발명의 CAPEAM*을 적용한 것보다 높은 GR 값(59.55)을 기록한 것을 확인할 수 있으나, 동일 조건에서 측정된 SR 값(42.64)과의 차이가 큰 것으로 보아, 복수의 서브-골 각각에 대한 실행 액션들의 수행 능력은 우수하지만, 복수의 서브-골 전체에 대한 실행 액션 전체의 수행 성공률은 월등히 떨어지는 것으로 확인된다. 반면에, 본 발명의 CAPEAM 모델은 우수한 SR 값을 기록한 것을 확인할 수 있다.

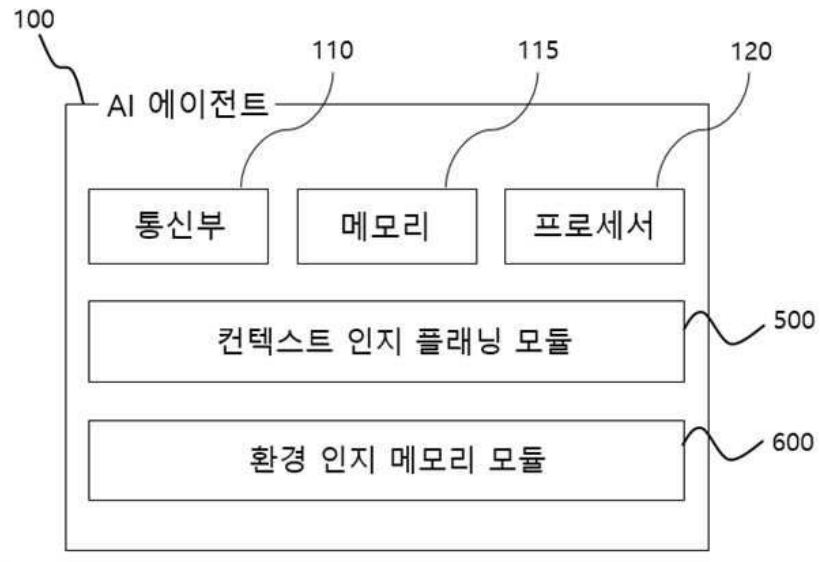
[0105] 이상 설명된 본 발명에 따른 실시예들은 다양한 컴퓨터 구성요소를 통하여 수행될 수 있는 프로그램 명령어의 형태로 구현되어 컴퓨터 판독 가능한 기록 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능한 기록 매체는 프로그램 명령어, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 컴퓨터 판독 가능한 기록 매체에 기록되는 프로그램 명령어는 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 분야의 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능한 기록 매체의 예에는, 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체, CD-ROM, DVD와 같은 광기록 매체, 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 ROM, RAM, 플래시 메모리 등과 같은 프로그램 명령어를 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령어의 예에는, 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드도 포함된다. 상기 하드웨어 장치는 본 발명에 따른 처리를 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

[0106] 이상에서 본 발명이 구체적인 구성요소 등과 같은 특정 사항들과 한정된 실시예 및 도면에 의해 설명되었으나, 이는 본 발명의 보다 전반적인 이해를 돕기 위해서 제공된 것일 뿐, 본 발명이 상기 실시예들에 한정되는 것은 아니며, 본 발명이 속하는 기술분야에서 통상적인 지식을 가진 자라면 이러한 기재로부터 다양한 수정 및 변형을 꾀할 수 있다.

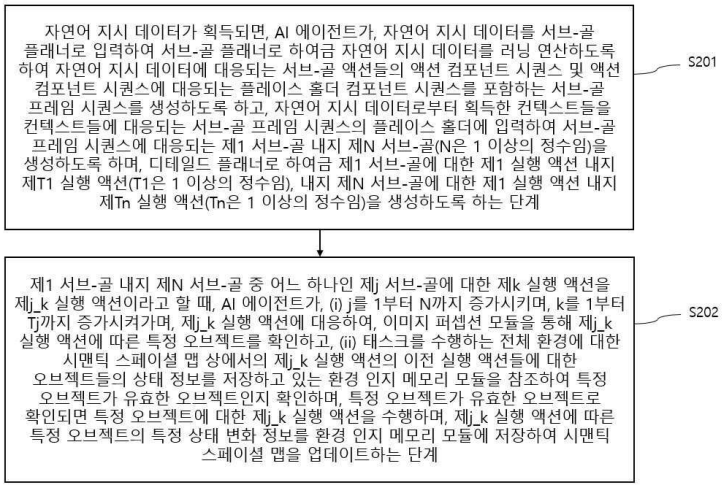
[0107] 따라서, 본 발명의 사상은 상기 설명된 실시예에 국한되어 정해져서는 아니 되며, 후술하는 특허청구범위뿐만 아니라 이 특허청구범위와 균등하게 또는 등가적으로 변형된 모든 것들은 본 발명의 사상의 범주에 속한다고 할 것이다.

도면

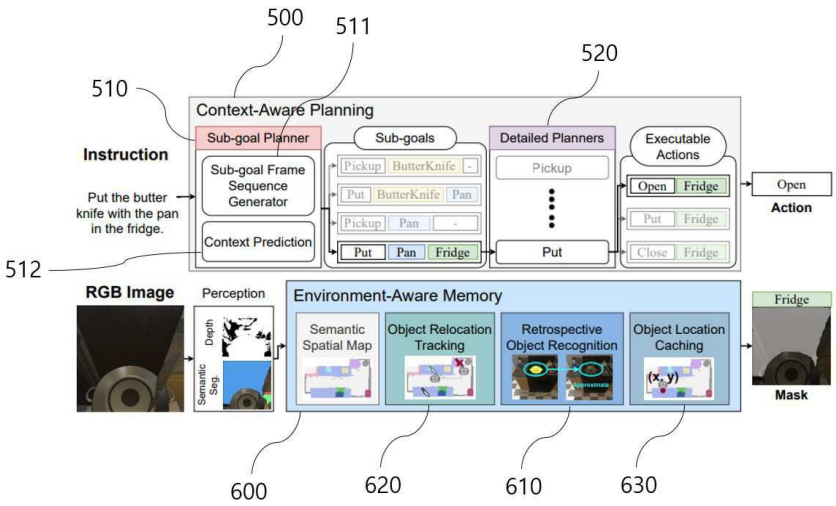
도면1



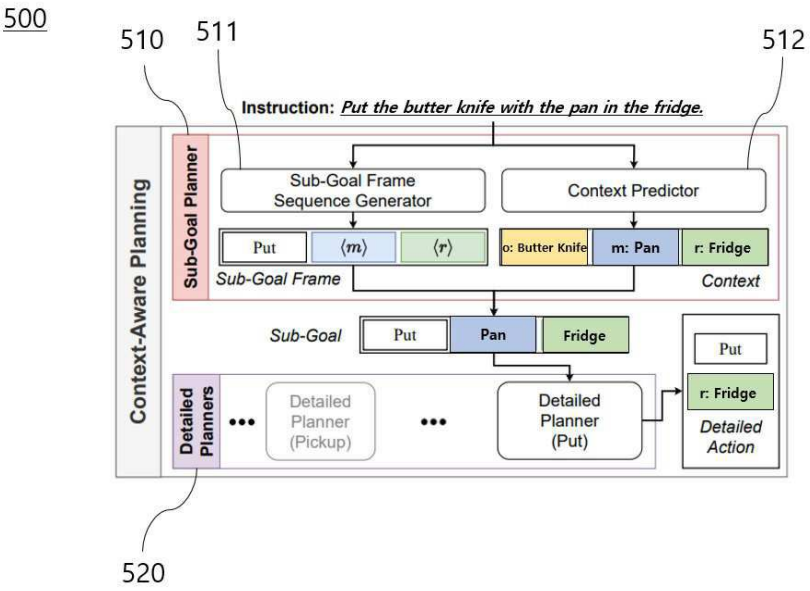
도면2


자연어 지시 데이터가 획득되면, AI 에이전트가, 자연어 지시 데이터를 서브-골 플래너로 입력하여 서브-골 플래너로 하여금 자연어 지시 데이터를 러닝 연산하도록 하여 자연어 지시 데이터에 대응되는 서브-골 액션들의 액션 컴포넌트 시퀀스 및 액션 컴포넌트 시퀀스에 대응되는 플레이스 홀더 컴포넌트 시퀀스를 포함하는 서브-골 프레임 시퀀스를 생성하도록 하고, 자연어 지시 데이터로부터 획득한 컨텍스트들을 컨텍스트들에 대응되는 서브-골 프레임 시퀀스의 플레이스 홀더에 입력하여 서브-골 프레임 시퀀스에 대응되는 제1 서브-골 내지 제N 서브-골(N은 1 이상의 정수임)을 생성하도록 하며, 디테일드 플래너로 하여금 제1 서브-골에 대한 제1 실행 액션 내지 제T1 실행 액션(T1은 1 이상의 정수임), 내지 제N 서브-골에 대한 제1 실행 액션 내지 제Tn 실행 액션(Tn은 1 이상의 정수임)을 생성하도록 하는 단계
S201
제1 서브-골 내지 제N 서브-골 중 어느 하나인 제j 서브-골에 대한 제k 실행 액션을 제j_k 실행 액션이라고 할 때, AI 에이전트가, (i) j를 1부터 N까지 증가시키며, k를 1부터 Tj까지 증가시켜가며, 제j_k 실행 액션에 대응하여, 이미지 퍼셉션 모듈을 통해 제j_k 실행 액션에 따른 특정 오브젝트를 확인하고, (ii) 태스크를 수행하는 전체 환경에 대한 시맨틱 스페이셜 맵 상에서의 제j_k 실행 액션의 이전 실행 액션들에 대한 오브젝트들의 상태 정보를 저장하고 있는 환경 인지 메모리 모듈을 참조하여 특정 오브젝트가 유효한 오브젝트인지 확인하며, 특정 오브젝트가 유효한 오브젝트로 확인되면 특정 오브젝트에 대한 제j_k 실행 액션을 수행하며, 제j_k 실행 액션에 따른 특정 오브젝트의 특정 상태 변화 정보를 환경 인지 메모리 모듈에 저장하여 시맨틱 스페이셜 맵을 업데이트하는 단계
S202

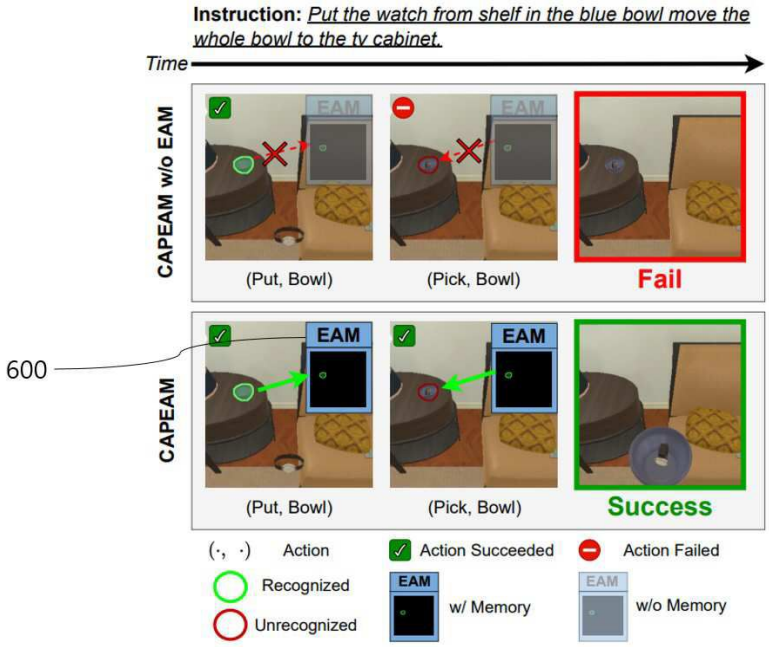
도면3


500
511
520
510
Context-Aware Planning
Instruction
Sub-goal Planner
Sub-goals
Detailed Planners
Executable Actions
Put the butter knife with the pan in the fridge.
Sub-goal Frame Sequence Generator
Context Prediction
Pickup ButterKnife -
Put ButterKnife Pan
Pickup Pan -
Put Pan Fridge
Pickup
Put
Open Fridge
Put Fridge
Close Fridge
Open
Action
512
RGB Image
Perception
Depth
Semantic Seg
Environment-Aware Memory
Semantic Spatial Map
Object Relocation Tracking
Retrospective Object Recognition
Object Location Caching
(x, y)
Fridge
Mask
600
620
610
630

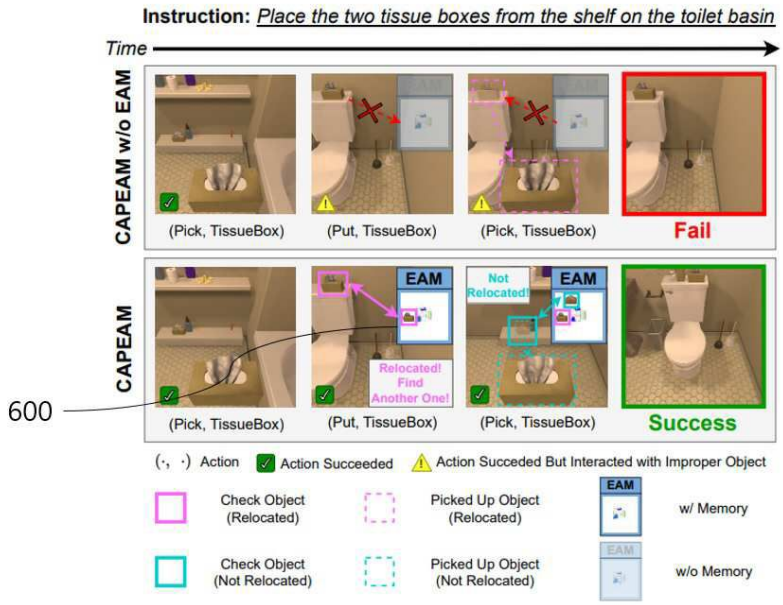
도면4


500
510
511
512
Instruction: Put the butter knife with the pan in the fridge.
Context-Aware Planning
Sub-Goal Planner
Sub-Goal Frame Sequence Generator
Context Predictor
Put
⟨m⟩
⟨r⟩
o: Butter Knife
m: Pan
r: Fridge
Sub-Goal Frame
Context
Sub-Goal
Put
Pan
Fridge
Detailed Planners
Detailed Planner (Pickup)
Detailed Planner (Put)
Put
r: Fridge
Detailed Action
520

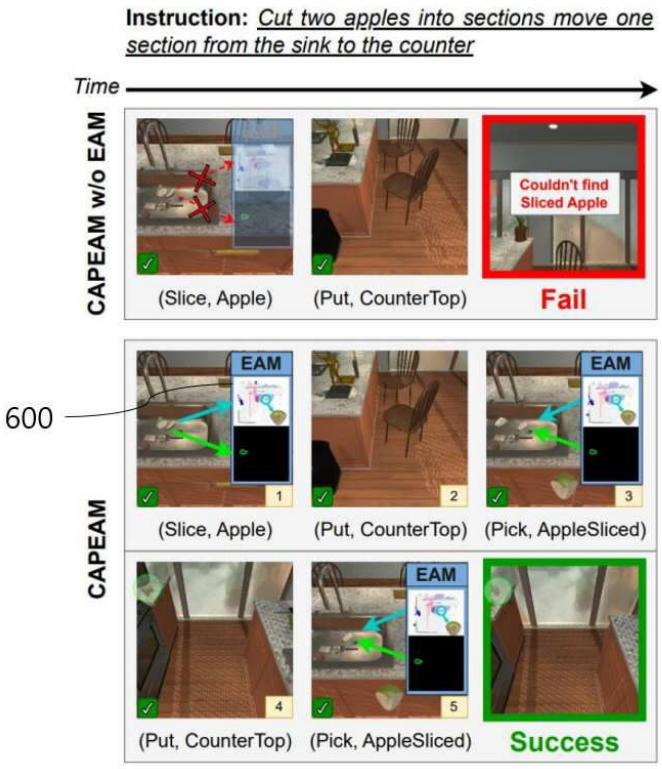
도면5a


Instruction: Put the watch from shelf in the blue bowl move the whole bowl to the tv cabinet.
Time
CAPEAM w/o EAM
EAM
(Put, Bowl)
(Pick, Bowl)
Fail
600
CAPEAM
EAM
(Put, Bowl)
(Pick, Bowl)
Success
(·, ·) Action
Action Succeeded
Action Failed
Recognized
Unrecognized
EAM
w/ Memory
EAM
w/o Memory

도면5b


Instruction: Place the two tissue boxes from the shelf on the toilet basin
Time
CAPEAM w/o EAM
(Pick, TissueBox)
(Put, TissueBox)
(Pick, TissueBox)
Fail
CAPEAM
EAM
Not Relocated!
Relocated! Find Another One!
(Pick, TissueBox)
(Put, TissueBox)
(Pick, TissueBox)
Success
600
(·, ·) Action
Action Succeeded
Action Succeeded But Interacted with Improper Object
Check Object (Relocated)
Picked Up Object (Relocated)
w/ Memory
Check Object (Not Relocated)
Picked Up Object (Not Relocated)
w/o Memory

도면5c


Instruction: Cut two apples into sections move one section from the sink to the counter
Time
CAPEAM w/o EAM
Couldn't find Sliced Apple
(Slice, Apple)
(Put, CounterTop)
Fail
CAPEAM
EAM
1
2
3
(Slice, Apple)
(Put, CounterTop)
(Pick, AppleSliced)
4
5
(Put, CounterTop)
(Pick, AppleSliced)
Success
600