

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국

(43) 국제공개일
2024년 5월 10일 (10.05.2024)



(10) 국제공개번호
WO 2024/096607 A1

- (51) 국제특허분류:
G05D 1/20 (2024.01) *G01C 21/34* (2006.01)
G06N 20/00 (2019.01)
- (21) 국제출원번호: PCT/KR2023/017342
- (22) 국제출원일: 2023년 11월 2일 (02.11.2023)
- (25) 출원언어: 한국어
- (26) 공개언어: 한국어
- (30) 우선권정보:
10-2022-0144397 2022년 11월 2일 (02.11.2022) KR
- (71) 출원인: 네이버랩스 주식회사 (NAVER LABS CORPORATION) [KR/KR]; 13561 경기도 성남시 분당구 정자일로 95, 18층, Gyeonggi-do (KR).
- (72) 발명자: 정승환 (CHUNG, Seunghwan); 13561 경기도 성남시 분당구 정자일로 95, NAVER 1784, Gyeonggi-do (KR). 최진영 (CHOI, Jinyoung); 13561 경기도

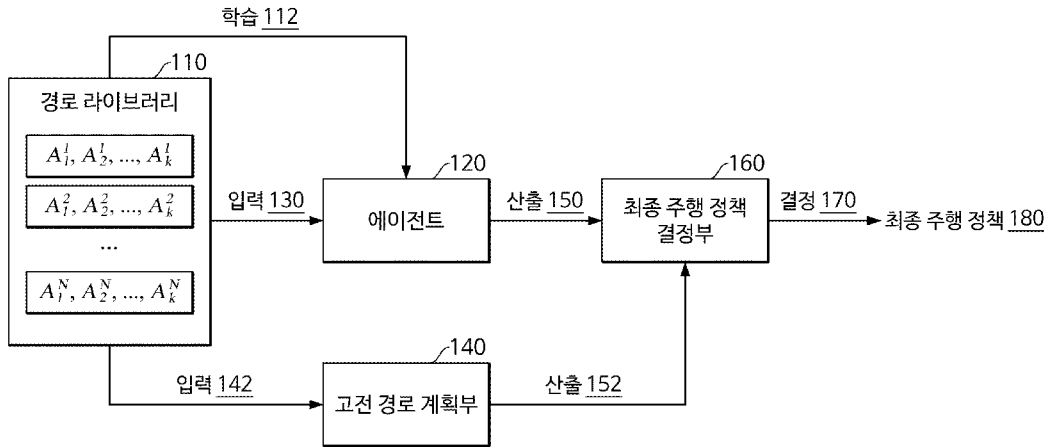
성남시 분당구 정자일로 95, NAVER 1784, Gyeonggi-do (KR). 황슬빈 (HWANG, Seulbin); 13561 경기도 성남시 분당구 정자일로 95, NAVER 1784, Gyeonggi-do (KR). 김건우 (KIM, Keonwoo); 13561 경기도 성남시 분당구 정자일로 95, NAVER 1784, Gyeonggi-do (KR). 변재정 (BYUN, Jaejung); 13561 경기도 성남시 분당구 정자일로 95, NAVER 1784, Gyeonggi-do (KR).

(74) 대리인: 김한솔 등 (KIM, Han Sol et al.); 06253 서울특별시 강남구 도곡로 117, 옥신타워 14층, AnK 특허법률사무소, Seoul (KR).

(81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO,

(54) Title: TRAVEL POLICY DETERMINATION METHOD AND SYSTEM BASED ON HYBRID TRAVEL ALGORITHM

(54) 발명의 명칭: 하이브리드 주행 알고리즘 기반 주행 정책 결정 방법 및 시스템



- 110 ... Trajectory library
112 ... Training
120 ... Agent
130, 142 ... Input
140 ... Classical trajectory planning unit
150, 152 ... Calculation
160 ... Final travel policy determination unit
170 ... Determination
180 ... Final travel policy

(57) Abstract: The present invention relates to a travel policy determination method based on a hybrid travel algorithm and performed by at least one processor. The travel policy determination method based on a hybrid travel algorithm comprises the steps of: obtaining an initial travel policy of an agent trained using a trajectory library; calculating a first numerical value related to one or more travel trajectories by a first method using the initial travel policy of the trained agent; calculating a second numerical value related to one or more travel trajectories by a second method; and determining a final travel policy of the agent on the basis of the calculated first numerical value and the calculated second numerical value.



NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW,
SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

- (84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의
역내 권리의 보호를 위하여): ARIPO (BW, CV, GH, GM,
KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ,
UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ,
TM), 유럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE,
ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC,
ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,
TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
KM, ML, MR, NE, SN, TD, TG).

공개:

— 국제조사보고서와 함께 (조약 제21조(3))

(57) 요약서: 본 개시는 적어도 하나의 프로세서에 의해 수행되는, 하이브리드 주행 알고리즘 기반 주행 정책 결정 방법에 관한 것이다. 하이브리드 주행 알고리즘 기반 주행 정책 결정 방법은, 경로 라이브러리를 이용하여 학습된 에이전트의 초기 주행 정책을 획득하는 단계, 학습된 에이전트의 초기 주행 정책을 이용한 제1 방식에 의하여, 하나 이상의 주행 경로와 관련된 제1 수치를 산출하는 단계, 제2 방식에 의하여, 하나 이상의 주행 경로와 관련된 제2 수치를 산출하는 단계 및 산출된 제1 수치 및 산출된 제2 수치에 기초하여, 에이전트의 최종 주행 정책을 결정하는 단계를 포함한다.

명세서

발명의 명칭: 하이브리드 주행 알고리즘 기반 주행 정책 결정 방법 및 시스템

기술분야

- [1] 본 개시는 하이브리드 주행 알고리즘 기반 주행 정책 결정 방법 및 시스템에 관한 것으로, 구체적으로, 학습된 에이전트의 초기 주행 정책 및 고전 경로 계획 알고리즘에 기초하여 하나 이상의 주행 경로와 관련된 수치를 산출함으로써, 최종 주행 정책을 결정하는 방법 및 시스템에 관한 것이다.

배경기술

- [2] 최근 4차 산업 혁명에 따른 자동화 기술로서 자율 주행 자동차 및 자율 이동 로봇에 대한 관심이 높아지고 있으며, 관련 기술에 대한 개발이 가속화되고 있다. 자율 주행 기술이 실현되기 위해서는 주변 지도 작성, 위치 인식 기술, 장애물 회피 기술뿐만 아니라 목적지까지의 경로를 결정하기 위한 경로 계획 기술이 필요하다.
- [3] 한편, 자율 주행을 위한 기존의 경로 계획 기술로서 강화학습 기반의 주행 알고리즘이 이용되고 있다. 그러나, 1-step 행동을 예측하는 강화학습 기반 주행 알고리즘의 경우, 동적 환경에서 회피 성능이 좋지만 전체 경로에 대한 예측이 불가능하고, 고전 경로 계획 알고리즘과 호환되지 않을 수 있다. 또한, multi-step 행동을 예측하는 강화학습 기반 주행 알고리즘의 경우에도, 전체 경로를 알 수 있으나 한 개의 경로만을 예측하기 때문에 고전 경로 계획 알고리즘과 호환되기 어려울 수 있다.

발명의 상세한 설명

기술적 과제

- [4] 본 개시는 상기와 같은 문제점을 해결하기 위한 하이브리드 주행 알고리즘 기반 주행 정책 결정 방법, 명령어들을 기록한 컴퓨터 판독가능한 비일시적 기록매체 및 장치(시스템)를 제공한다.

과제 해결 수단

- [5] 본 개시는 방법, 장치(시스템) 또는 판독 가능 저장 매체에 저장된 컴퓨터 프로그램을 포함한 다양한 방식으로 구현될 수 있다.
- [6] 본 개시의 일 실시예에 따르면, 적어도 하나의 프로세서에 의해 수행되는, 하이브리드(hybrid) 주행 알고리즘 기반 주행 정책(policy) 결정 방법은, 경로 라이브러리(trjectory library)를 이용하여 학습된 에이전트의 초기 주행 정책을 획득하는 단계, 학습된 에이전트의 초기 주행 정책을 이용한 제1 방식에 의하여, 하나 이상의 주행 경로와 관련된 제1 수치를 산출하는 단계, 제2 방식에 의하여, 하나 이상의 주행 경로와 관련된 제2 수치를 산출하는 단계 및 산출된 제1 수치 및 산출된 제2 수치에 기초하여, 최종 주행 정책을 결정하는 단계를 포함한다.

- [7] 본 개시의 일 실시예에 따른 방법을 컴퓨터에서 실행하기 위한 명령어들을 기록한 컴퓨터로 판독 가능한 비일시적 기록 매체가 제공된다.
- [8] 본 개시의 일 실시예에 따른, 정보 처리 시스템은, 통신 모듈, 메모리, 메모리와 연결되고, 메모리에 포함된 컴퓨터 판독 가능한 적어도 하나의 프로그램을 실행하도록 구성된 적어도 하나의 프로세서를 포함하고, 경로 라이브러리를 이용하여 학습된 에이전트의 초기 주행 정책을 획득하고, 학습된 에이전트의 초기 주행 정책을 이용한 제1 방식에 의하여, 하나 이상의 주행 경로와 관련된 제1 수치를 산출하고, 제2 방식에 의하여, 하나 이상의 주행 경로와 관련된 제2 수치를 산출하고, 산출된 제1 수치 및 산출된 제2 수치에 기초하여, 최종 주행 정책을 결정하기 위한 명령어들을 포함한다.

발명의 효과

- [9] 본 개시의 일 실시예에 따르면, 강화 학습 기반 알고리즘으로부터 출력된 주행 경로가, 하나의 주행 경로가 아닌 복수의 경로들을 포함하고 있기 때문에, 고전 경로 계획 알고리즘과 호환될 수 있다.
- [10] 본 개시의 일 실시예에 따르면, 강화 학습 기반의 에이전트의 초기 주행 정책이 경로 라이브러리로부터 가장 높은 보상(reward)을 주는 주행 경로를 선택하도록 학습됨으로써, 선택된 주행 경로를 통해 에이전트가 현재 어떤 경로로 주행할 것인지 예측할 수 있다.
- [11] 본 개시의 일 실시예에 따르면, 강화학습 기반의 주행 알고리즘과 고전 경로 계획 알고리즘을 결합한 하이브리드 주행 알고리즘을 이용함으로써, 동적인 환경에서 대상 객체(예: 로봇 등)의 회피 성능이 향상될 수 있다.
- [12] 본 개시의 효과는 이상에서 언급한 효과로 제한되지 않으며, 언급되지 않은 다른 효과들은 청구범위의 기재로부터 본 개시가 속하는 기술분야에서 통상의 지식을 가진 자(“통상의 기술자”라 함)에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

- [13] 본 개시의 실시예들은, 이하 설명하는 첨부 도면들을 참조하여 설명될 것이며, 여기서 유사한 참조 번호는 유사한 요소들을 나타내지만, 이에 한정되지는 않는다.
- [14] 도 1은 본 개시의 일 실시예에 따른 하이브리드 주행 알고리즘 기반 주행 정책 결정 방법의 일 예시를 나타내는 도면이다.
- [15] 도 2는 본 개시의 일 실시예에 따른 정보 처리 시스템이 복수의 대상 객체인 자율 이동 로봇과 통신 가능하도록 연결된 구성을 나타내는 개요도이다.
- [16] 도 3는 본 개시의 일 실시예에 따른 하이브리드 주행 알고리즘 기반 주행 정책 결정 방법을 제공하는 정보 처리 시스템을 나타내는 블록도이다.
- [17] 도 4는 본 개시의 일 실시예에 따른 프로세서의 내부 구성을 나타내는 도면이다.
- [18] 도 5는 본 개시의 일 실시예에 따른 강화학습 모델을 나타내는 예시도이다.

- [19] 도 6은 본 개시의 일 실시예에 따른 경로 라이브러리, 에이전트 및 환경 사이의 상호 작용에 대한 예시는 나타내는 도면이다.
- [20] 도 7은 본 개시의 일 실시예에 따라 경로 라이브러리가 업데이트되는 예시를 나타내는 도면이다.
- [21] 도 8은 본 개시의 일 실시예에 따라 에이전트의 초기 주행 정책이 학습되는 예시를 나타내는 도면이다.
- [22] 도 9은 본 개시의 다른 실시예에 따라 에이전트의 초기 주행 정책이 학습되는 예시를 나타내는 도면이다.
- [23] 도 10은 본 개시의 일 실시예에 따라 주행 경로에 대한 수치를 산출하고, 최종 주행 정책을 결정하는 예시를 나타내는 도면이다.
- [24] 도 11은 본 개시의 일 실시예에 따른 하이브리드 주행 알고리즘 기반 주행 정책 결정 방법의 예시를 나타내는 흐름도이다.

발명의 실시를 위한 형태

- [25] 이하, 본 개시의 실시를 위한 구체적인 내용을 첨부된 도면을 참조하여 상세히 설명한다. 다만, 이하의 설명에서는 본 개시의 요지를 불필요하게 흐릴 우려가 있는 경우, 널리 알려진 기능이나 구성에 관한 구체적 설명은 생략하기로 한다.
- [26] 첨부된 도면에서, 동일하거나 대응하는 구성요소에는 동일한 참조부호가 부여되어 있다. 또한, 이하의 실시예들의 설명에 있어서, 동일하거나 대응되는 구성요소를 중복하여 기술하는 것이 생략될 수 있다. 그러나, 구성요소에 관한 기술이 생략되어도, 그러한 구성요소가 어떤 실시예에 포함되지 않는 것으로 의도되지 않는다는 점에 유의한다.
- [27] 개시된 실시예의 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 후술되어 있는 실시예들을 참조하면 명확해질 것이다. 그러나, 본 개시는 이하에서 개시되는 실시예들에 한정되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있으며, 단지 본 실시예들은 본 개시가 완전하도록 하고, 본 개시가 통상의 기술자에게 발명의 범주를 완전하게 알려주기 위해 제공되는 것일 뿐이다.
- [28] 본 명세서에서 사용되는 용어에 대해 간략히 설명하고, 개시된 실시예에 대해 구체적으로 설명하기로 한다. 본 명세서에서 사용되는 용어는 본 개시에서의 기능을 고려하면서 가능한 현재 널리 사용되는 일반적인 용어들을 선택하였으나, 이는 관련 분야에 종사하는 기술자의 의도 또는 판례, 새로운 기술의 출현 등에 따라 달라질 수 있다. 또한, 특정한 경우는 출원인이 임의로 선정한 용어도 있으며, 이 경우 해당되는 발명의 설명 부분에서 상세히 그 의미를 기재할 것이다. 따라서, 본 개시에서 사용되는 용어는 단순한 용어의 명칭이 아닌, 그 용어가 가지는 의미와 본 개시의 전반에 걸친 내용을 토대로 정의되어야 한다.
- [29] 본 명세서에서의 단수의 표현은 문맥상 명백하게 단수인 것으로 특정하지 않는 한, 복수의 표현을 포함한다. 또한, 복수의 표현은 문맥상 명백하게 복수인 것으로

로 특정하지 않는 한, 단수의 표현을 포함한다. 명세서 전체에서 어떤 부분이 어떤 구성요소를 포함한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있음을 의미한다.

[30] 또한, 명세서에서 사용되는 '모듈' 또는 '부'라는 용어는 소프트웨어 또는 하드웨어 구성요소를 의미하며, '모듈' 또는 '부'는 어떤 역할들을 수행한다. 그렇지만, '모듈' 또는 '부'는 소프트웨어 또는 하드웨어에 한정되는 의미는 아니다. '모듈' 또는 '부'는 어드레싱할 수 있는 저장 매체에 있도록 구성될 수도 있고 하나 또는 그 이상의 프로세서들을 재생시키도록 구성될 수도 있다. 따라서, 일 예로서, '모듈' 또는 '부'는 소프트웨어 구성요소들, 객체지향 소프트웨어 구성요소들, 클래스 구성요소들 및 태스크 구성요소들과 같은 구성요소들과, 프로세스들, 함수들, 속성들, 프로시저들, 서브루틴들, 프로그램 코드의 세그먼트들, 드라이버들, 펌웨어, 마이크로 코드, 회로, 데이터, 데이터베이스, 데이터 구조들, 테이블들, 어레이들 또는 변수들 중 적어도 하나를 포함할 수 있다. 구성요소들과 '모듈' 또는 '부'들은 안에서 제공되는 기능은 더 작은 수의 구성요소들 및 '모듈' 또는 '부'들로 결합되거나 추가적인 구성요소들과 '모듈' 또는 '부'들로 더 분리될 수 있다.

[31] 본 개시의 일 실시예에 따르면, '모듈' 또는 '부'는 프로세서 및 메모리로 구현될 수 있다. '프로세서'는 범용 프로세서, 중앙 처리 장치(CPU), 마이크로프로세서, 디지털 신호 프로세서(DSP), 제어기, 마이크로제어기, 상태 머신 등을 포함하도록 넓게 해석되어야 한다. 몇몇 환경에서, '프로세서'는 주문형 반도체(ASIC), 프로그램가능 로직 디바이스(PLD), 필드 프로그램가능 게이트 어레이(FPGA) 등을 지칭할 수도 있다. '프로세서'는, 예를 들어, DSP와 마이크로프로세서의 조합, 복수의 마이크로프로세서들의 조합, DSP 코어와 결합한 하나 이상의 마이크로프로세서들의 조합, 또는 임의의 다른 그러한 구성들의 조합과 같은 처리 디바이스들의 조합을 지칭할 수도 있다. 또한, '메모리'는 전자 정보를 저장 가능한 임의의 전자 컴포넌트를 포함하도록 넓게 해석되어야 한다. '메모리'는 임의 액세스 메모리(RAM), 판독-전용 메모리(ROM), 비-휘발성 임의 액세스 메모리(NVRAM), 프로그램가능 판독-전용 메모리(PROM), 소거-프로그램가능 판독 전용 메모리(EPROM), 전기적으로 소거가능 PROM(EEPROM), 플래쉬 메모리, 자기 또는 광학 데이터 저장장치, 레지스터들 등과 같은 프로세서-판독가능 매체의 다양한 유형들을 지칭할 수도 있다. 프로세서가 메모리로부터 정보를 판독하고/하거나 메모리에 정보를 기록할 수 있다면 메모리는 프로세서와 전자 통신 상태에 있다고 불린다. 프로세서에 집적된 메모리는 프로세서와 전자 통신 상태에 있다.

[32] 본 개시에서, '시스템'은 서버 장치와 클라우드 장치 중 적어도 하나의 장치를 포함할 수 있으나, 이에 한정되는 것은 아니다. 예를 들어, 시스템은 하나 이상의 서버 장치로 구성될 수 있다. 다른 예로서, 시스템은 하나 이상의 클라우드 장치로 구성될 수 있다. 또 다른 예로서, 시스템은 서버 장치와 클라우드 장치가 함께 구성되어 동작될 수 있다.

- [33] 본 개시에서, '복수의 A의 각각' 또는 '복수의 A 각각'은 복수의 A에 포함된 모든 구성 요소의 각각을 지칭하거나, 복수의 A에 포함된 일부 구성 요소의 각각을 지칭할 수 있다. 예를 들어, 복수의 주행 경로의 각각은, 복수의 주행 경로에 포함된 모든 주행 경로의 각각을 지칭하거나 복수의 주행 경로에 포함된 일부 주행 경로의 각각을 지칭할 수 있다.
- [34] 본 개시에서, 강화학습(Reinforcement Learning, RL)은 기계학습의 한 영역으로, 컴퓨터가 주어진 상태(state)에 대해 선택 가능한 액션(action)들 중 보상(reward)을 최대화하는 액션을 선택하는 학습 방법을 지칭할 수 있다. 여기서, 강화학습의 대상이 되는 컴퓨터 프로그램을 에이전트(agent)라고 지칭할 수 있고, 에이전트는 주어진 환경을 탐색하여 현재의 상태에서 자신이 취할 수 있는 액션의 확률을 나타내는 정책(policy)을 수립할 수 있다. 또한, 주어진 환경(environment)에 의해 에이전트에게 특정 상태(state)가 주어지면(에이전트가 관찰에 의해 상태를 얻게 된다고 볼 수도 있음), 에이전트는 상태에 따라 액션을 하고, 환경은 에이전트에게 보상을 주게 된다. 이러한 과정으로 에이전트는 환경과 상호작용을 하며 보상을 많이 취할 수 있는 액션들 또는 정책을 학습할 수 있다. 즉, 주어진 환경과 에이전트 사이에서 상태, 액션, 보상을 상호작용하면서 에이전트 또는 에이전트의 정책이 학습되고, 이를 통해 에이전트가 최대의 보상을 받을 수 있는 정책을 수립하는 것이 강화학습의 목표일 수 있다.
- [35] 본 개시에서, '기계학습 모델'은 주어진 입력에 대한 해답(answer)을 추론하는데 사용하는 임의의 모델을 포함할 수 있다. 일 실시예에 따르면, 기계학습 모델은 입력 레이어(층), 복수 개의 은닉 레이어 및 출력 레이어를 포함한 인공신경망 모델을 포함할 수 있다. 여기서, 각 레이어는 복수의 노드를 포함할 수 있다. 본 개시에서, 제1 기계학습 모델, 제2 기계학습 모델, 제3 기계학습 모델 등 복수의 기계학습 모델을 별도의 기계학습 모델로서 설명하나, 이에 한정되지 않으며, 복수의 기계학습 모델의 일부 또는 전체는 하나의 기계학습 모델로 구현될 수 있다. 또한, 본 개시에서, 기계학습 모델은 인공신경망 모델을 지칭할 수 있으며, 인공신경망 모델은 기계학습 모델을 지칭할 수 있다.
- [36] 도 1은 본 개시의 일 실시예에 따른 하이브리드 주행 알고리즘 기반 주행 정책 결정 방법의 일 예시를 나타내는 도면이다. 일 실시예에 따르면, 경로 라이브러리(110)를 이용하여 에이전트(120)가 학습될 수 있다(112). 이 경우, 경로 라이브러리(110)는 대상 객체가 주행할 수 있는 복수의 주행 경로를 포함하도록 구축될 수 있으며, 전문가 에이전트(expert agent)의 숙련된 주행을 위해 사용하는 주행 경로를 포함하도록 업데이트될 수 있다. 그 후, 학습된 에이전트(120)의 초기 주행 정책이 획득될 수 있다. 이 경우, 에이전트(120)의 초기 주행 정책은, 경로 라이브러리(110)에 포함된 복수의 주행 경로 중에서, 가장 높은 보상을 주는 주행 경로를 선택하도록 학습된 임의의 알고리즘을 포함할 수 있으며, 예를 들어, 기계학습 모델(예: 인공신경망 모델 등)을 이용하여 학습될 수 있다. 에이전트(120)의 초기 주행 정책에 대한 세부적인 내용을 도 7 및 도 8에서 후술한다.

- [37] 일 실시예에 따르면, 학습된 에이전트(120)는 경로 라이브러리(110)에 포함된 하나 이상의 주행 경로를 입력받고(130), 초기 주행 정책을 이용하여, 주행 경로와 관련된 제1 수치를 산출할 수 있다(112). 이 경우, 주행 경로와 관련된 제1 수치는 하나 이상의 주행 경로에 대한 가치(value)를 포함할 수 있다.
- [38] 일 실시예에 따르면, 고전 경로 계획부(140)는 경로 라이브러리(110)에 포함된 하나 이상의 주행 경로를 입력받아(142) 고전 경로 계획 알고리즘을 이용하여, 하나 이상의 주행 경로와 관련된 제2 수치를 산출(152)할 수 있다. 이 경우, 주행 경로와 관련된 제2 수치는 하나 이상의 주행 경로에 대한 비용(cost)을 포함할 수 있다.
- [39] 그 후, 최종 주행 정책 결정부(160)는 에이전트(120)에 의해 산출된 수치 및 고전 경로 계획부(140)에 의해 산출된 수치를 수신하고, 이를 기초로 에이전트의 최종 주행 정책(180)을 결정할 수 있다(170). 예를 들어, 최종 주행 정책 결정부는 산출된 수치 각각에 가중치를 적용하여 합산한 수치가 가장 큰 주행 경로를 선택하도록 최종 주행 정책(180)을 결정할 수 있다. 이러한 가중치는 산출된 수치 각각의 비중을 조절하기 위해 동적으로 조절될 수 있다.
- [40] 이와 같이, 강화학습 기반의 주행 알고리즘을 이용한 에이전트(120) 및 고전 경로 계획 알고리즘을 이용한 고전 경로 계획부(140)로부터 산출된 수치들을 이용함으로써, 동적인 환경에서 회피 성능을 향상시킬 수 있는 주행 정책이 결정되거나 설립될 수 있다.
- [41] 도 2는 본 개시의 일 실시예에 따른 정보 처리 시스템(230)이 복수의 대상 객체인 자율 이동 로봇(210_1, 210_2, 210_3)과 통신 가능하도록 연결된 구성을 나타내는 개요도이다. 정보 처리 시스템(230)은 주행 정책을 결정하는 시스템(들)을 포함할 수 있다. 일 실시예에서, 정보 처리 시스템(230)은 주행 정책을 결정하는 것과 관련된 컴퓨터 실행 가능한 프로그램(예를 들어, 다운로드 가능한 어플리케이션) 및 데이터를 저장, 제공 및 실행할 수 있는 하나 이상의 서버 장치 및/또는 데이터베이스, 또는 클라우드 컴퓨팅 서비스 기반의 하나 이상의 분산 컴퓨팅 장치 및/또는 분산 데이터베이스를 포함할 수 있다. 정보 처리 시스템(230)은 어플리케이션(예: 로봇 어플리케이션 및/또는 인공지능 어플리케이션 등)을 통해 입력되는 신호에 대응하는 정보를 제공하거나 대응하는 처리를 수행할 수 있다. 예를 들어, 정보 처리 시스템(230)은 주행 정책 결정 및/또는 주행 경로 예측과 관련된 임의의 어플리케이션을 통해 복수의 자율 이동 로봇(210_1, 210_2, 210_3)을 제어할 수 있다.
- [42] 정보 처리 시스템(230)은 네트워크(220)를 통해 복수의 자율 이동 로봇(210_1, 210_2, 210_3)과 통신할 수 있다. 네트워크(220)는 복수의 자율 이동 로봇(210_1, 210_2, 210_3)과 정보 처리 시스템(230) 사이의 통신이 가능하도록 구성될 수 있다. 네트워크(220)는 설치 환경에 따라, 예를 들어, 이더넷(Ethernet), 유선 홈 네트워크(Power Line Communication), 전화선 통신 장치 및 RS-serial 통신 등의 유선 네트워크, 이동통신망, WLAN(Wireless LAN), Wi-Fi, Bluetooth 및 ZigBee 등과 같

은 무선 네트워크 또는 그 조합으로 구성될 수 있다. 통신 방식은 제한되지 않으며, 네트워크(220)가 포함할 수 있는 통신망(일례로, 이동통신망, 유선 인터넷, 무선 인터넷, 방송망, 위성망 등)을 활용하는 통신 방식뿐만 아니라 자율 이동 로봇(210_1, 210_2, 210_3) 사이의 근거리 무선 통신 역시 포함될 수 있다.

[43] 도 2에서 전문 서비스용 자율 이동 로봇(210_1), 산업용 자율 이동 로봇(210_2) 및 개인 서비스용 자율 이동 로봇(210_3)이 로봇의 예로서 도시되었으나, 이에 한정되지 않으며, 자율 이동 로봇(210_1, 210_2, 210_3)은 유선 및/또는 무선 통신이 가능하고 로봇 어플리케이션, 주행 정책 결정 어플리케이션, 인공지능 어플리케이션 등이 설치되어 실행될 수 있는 임의의 컴퓨팅 장치일 수 있다. 예를 들어, 자율 이동 로봇(210_1, 210_2, 210_3)은 자율 주행 차량 등을 포함할 수 있다. 또한, 도 2에는 3개의 자율 이동 로봇(210_1, 210_2, 210_3)이 네트워크(220)를 통해 정보 처리 시스템(230)과 통신하는 것으로 도시되어 있으나, 이에 한정되지 않으며, 상이한 수의 장치가 네트워크(220)를 통해 정보 처리 시스템(230)과 통신하도록 구성될 수도 있다.

[44] 일 실시예에서, 자율 이동 로봇(210_1, 210_2, 210_3)은 네트워크(220)를 통해 주행 정책과 관련된 데이터 요청을 정보 처리 시스템(230)으로 전송하고, 정보 처리 시스템(230)으로부터 주행 정책 및/또는 주행 경로와 관련된 데이터를 수신할 수 있다. 자율 이동 로봇(210_1, 210_2, 210_3)이 주행 정책 및/또는 주행 경로와 관련된 데이터를 수신하는 것에 응답하여, 주행 정책 및/또는 주행 경로에 따라 자율적으로 이동 또는 운행될 수 있다.

[45] 도 3는 본 개시의 일 실시예에 따른 하이브리드 주행 알고리즘 기반 주행 정책 결정 방법을 제공하는 정보 처리 시스템(230)을 나타내는 블록도이다. 주행 정책 결정 위한 정보 처리 시스템(230)은 메모리(310), 프로세서(320), 통신 모듈(330) 및 입출력 인터페이스(340)를 포함할 수 있다. 도 3에 도시된 바와 같이, 정보 처리 시스템(230)은 통신 모듈(330)을 이용하여 네트워크를 통해 정보 및/또는 데이터를 통신할 수 있도록 구성될 수 있다.

[46] 메모리(310)는 비-일시적인 임의의 컴퓨터 판독 가능한 기록매체를 포함할 수 있다. 일 실시예에 따르면, 메모리(310)는 디스크 드라이브, SSD(solid state drive), 플래시 메모리(flash memory) 등과 같은 비소멸성 대용량 저장 장치(permanent mass storage device)를 포함할 수 있다. 다른 예로서, ROM, SSD, 플래시 메모리, 디스크 드라이브 등과 같은 비소멸성 대용량 저장 장치는 메모리와는 구분되는 별도의 영구 저장 장치로서 정보 처리 시스템(2300)에 포함될 수 있다. 또한, 메모리(310)에는 운영체제와 적어도 하나의 프로그램 코드(예를 들어, 정보 처리 시스템(2300)에 설치되어 구동되는 초기 주행 정책 획득 명령, 하나 이상의 주행 경로와 관련된 수치 산출 명령, 최종 주행 정책 결정 명령 등)가 저장될 수 있다. 도 3에서, 메모리(310)는 단일 메모리인 것으로 도시되었지만, 이는 설명의 편의를 위한 것일 뿐이며, 메모리(310)는 복수의 메모리 및/또는 버퍼 메모리를 포함할 수 있다.

- [47] 이러한 소프트웨어 구성요소들은 메모리(310)와는 별도의 컴퓨터에서 판독 가능한 기록매체로부터 로딩될 수 있다. 이러한 별도의 컴퓨터에서 판독 가능한 기록매체는 이러한 정보 처리 시스템(230)에 직접 연결가능한 기록 매체를 포함할 수 있는데, 예를 들어, 플로피 드라이브, 디스크, 테이프, DVD/CD-ROM 드라이브, 메모리 카드 등의 컴퓨터에서 판독 가능한 기록매체를 포함할 수 있다. 다른 예로서, 소프트웨어 구성요소들은 컴퓨터에서 판독 가능한 기록매체가 아닌 통신 모듈(330)을 통해 메모리(310)에 로딩될 수도 있다. 예를 들어, 적어도 하나의 프로그램은 개발자들 또는 어플리케이션의 설치 파일을 배포하는 파일 배포 시스템이 통신 모듈(330)을 통해 제공하는 파일들에 의해 설치되는 컴퓨터 프로그램(예를 들어, 주행 정책 결정을 위한 프로그램 등)에 기반하여 메모리(310)에 로딩될 수 있다.
- [48] 프로세서(320)는 기본적인 산술, 로직 및 입출력 연산을 수행함으로써, 컴퓨터 프로그램의 명령을 처리하도록 구성될 수 있다. 명령은 메모리(310) 또는 통신 모듈(330)에 의해 사용자 단말(미도시) 또는 다른 외부 시스템으로 제공될 수 있다. 예를 들어, 프로세서(320)는 경로 라이브러리를 이용하여 학습된 에이전트의 초기 주행 정책을 획득하거나 학습시킬 수 있다.
- [49] 통신 모듈(330)은 네트워크를 통해 사용자 단말(미도시)과 정보 처리 시스템(230)이 서로 통신하기 위한 구성 또는 기능을 제공할 수 있으며, 정보 처리 시스템(230)이 외부 시스템(일례로 별도의 클라우드 시스템 등)과 통신하기 위한 구성 또는 기능을 제공할 수 있다. 일례로, 정보 처리 시스템(2300)의 프로세서(320)의 제어에 따라 제공되는 제어 신호, 명령, 데이터 등이 통신 모듈(330)과 네트워크를 거쳐 사용자 단말 및/또는 외부 시스템의 통신 모듈을 통해 사용자 단말 및/또는 외부 시스템으로 전송될 수 있다. 예를 들어, 외부 시스템(예: 주행 경로 결정 시스템)은 정보 처리 시스템(230)으로부터 주행 정책 정보 등을 전달받을 수 있다.
- [50] 또한, 정보 처리 시스템(230)의 입출력 인터페이스(340)는 정보 처리 시스템(230)과 연결되거나 정보 처리 시스템(230)이 포함할 수 있는 입력 또는 출력을 위한 장치(미도시)와의 인터페이스를 위한 수단일 수 있다. 예를 들면, 입출력 인터페이스(340)는 PCI express 인터페이스, 이더넷(ethernet) 인터페이스 중 적어도 하나를 포함할 수 있다. 도 3에서는 입출력 인터페이스(340)가 프로세서(320)와 별도로 구성된 요소로서 도시되었으나, 이에 한정되지 않으며, 입출력 인터페이스(340)가 프로세서(320)에 포함되도록 구성될 수 있다. 정보 처리 시스템(230)은 도 3의 구성요소들보다 더 많은 구성요소들을 포함할 수 있다. 그러나, 대부분의 종래기술적 구성요소들을 명확하게 도시할 필요성은 없다.
- [51] 정보 처리 시스템(230)의 프로세서(320)는 복수의 사용자 단말 및/또는 복수의 외부 시스템으로부터 수신된 정보 및/또는 데이터를 관리, 처리 및/또는 저장하도록 구성될 수 있다. 일 실시예에 따르면, 프로세서(320)는 주행 경로 데이터를 수신할 수 있다. 프로세서는 수신된 주행 경로 데이터에 기초하여 경로 라이브러

리를 업데이트 할 수 있다. 도 3에서, 프로세서(320)는 단일 프로세서인 것으로 도시되었지만, 이는 설명의 편의를 위한 것일 뿐이며, 프로세서(320)는 복수의 프로세서를 포함할 수 있다.

- [52] 도 4는 본 개시의 일 실시예에 따른 프로세서(320)의 내부 구성을 나타내는 도면이다. 프로세서(320)는 경로 라이브러리(110), 학습부(430), 에이전트(120), 고전 경로 계획부(140) 및 최종 주행 정책 결정부(160)를 포함할 수 있다. 이 경우, 경로 라이브러리(110), 에이전트(120), 학습부(430), 고전 경로 계획부(140) 및 최종 주행 정책 결정부(160)는 임의의 저장 매체(예: 메모리(310) 등)에 저장되거나 로딩되어, 프로세서(320)에 의해 접근되거나 실행될 수 있다. 도 4에서는 프로세서(320)의 각각의 구성 요소들은 기능적으로 구분되는 기능 요소들을 나타낸 것으로, 복수의 구성 요소가 실제 물리적 환경에서 서로 통합되는 형태로 구현될 수 있다. 또한, 도 4에서는 프로세서(320)는 경로 라이브러리(110), 에이전트(120), 학습부(430), 고전 경로 계획부(140) 및 최종 주행 정책 결정부(160)로 나누어서 구현되었으나, 이에 한정되지 않으며, 일부 구성이 생략되거나 다른 구성이 추가될 수 있다.
- [53] 일 실시예에 따르면, 경로 라이브러리(110)는 대상 객체가 주행할 수 있는 복수의 주행 경로를 포함하도록 구축될 수 있다. 예를 들어, 경로 라이브러리(110)는 대상 지점까지의 도달할 수 있는 주행 경로를 단계 별로 분류하여 저장할 수 있다. 이렇게 저장된 주행 경로는 저장 매체(예: 메모리(310), 버퍼 메모리, 외부 저장 매체 등)에 이동하여 저장되거나, 학습부(430), 에이전트(120) 및/또는 고전 경로 계획부(140)에 제공될 수 있다.
- [54] 경로 라이브러리(110)는 전문가 에이전트(expert agent)의 숙련된 주행을 위해 필요한 주행 경로를 구축하도록 업데이트될 수 있다. 이 경우, 대상 객체(예: 로봇 등)의 상태와 액션을 입력으로 하여 기대 보상을 출력하도록 구성된 기계학습 모델이 이용될 수 있다. 경로 라이브러리의 업데이트와 관련된 세부적인 내용은 도 7에서 후술한다.
- [55] 일 실시예에 따르면, 학습부(430)는 경로 라이브러리를 이용하여 에이전트(120)를 학습시킬 수 있다. 예를 들어, 상기 경로 라이브러리에 포함된 복수의 주행 경로 중에서, 가장 높은 보상을 주는 주행 경로를 선택하도록 에이전트(120)를 학습시킬 수 있다. 이를 통해, 대상 객체가 목적지에 충돌하지 않고 도달하도록 하는 주행 경로를 에이전트(120)가 선택하도록 하는 초기 주행 정책을 획득할 수 있다.
- [56] 또한, 일 실시예에 따르면, 학습부(430)는 경로 라이브러리(110)에 포함된 주행 경로를 업데이트할 수 있다. 예를 들어, 학습부(430)는 전문가 에이전트의 주행 경로를 포함하도록 경로 라이브러리를 업데이트할 수 있다. 이 경우, 학습부(430)는 경로 라이브러리를 업데이트하는데 있어 프로세서에 의해 접근 가능한 임의의 저장매체에 저장된 기계학습 모델을 이용할 수 있다.

- [57] 일 실시예에 따르면, 에이전트(120)는 현재 어떤 경로로 주행할 것인지 주행 경로를 선택할 수 있다. 이렇게 선택된 주행 경로는 대상 객체의 이동 경로로 결정될 수 있다. 일 실시예에 따르면, 에이전트(120)는 경로 라이브러리(110)를 이용하여 학습됨으로써, 초기 주행 정책이 학습되거나 설립될 수 있다. 예를 들어, 에이전트(120)는 경로 라이브러리 중 가장 높은 보상을 받을 수 있는 경로를 선택하도록 학습되며, 선택된 주행 경로를 통해 에이전트(120)가 현재 어떤 경로로 주행할 것인지 결정할 수 있다.
- [58] 또한, 일 실시예에 따르면, 에이전트(120)는 주행 정책을 이용하여 하나 이상의 주행 경로와 관련된 제1 수치를 산출할 수 있다. 예를 들어, 에이전트(120)는 경로 라이브러리(110)를 이용하여 설립된 초기 주행 정책을 통해 경로 라이브러리(110)에 포함된 복수의 주행 경로의 각각에 대한 가치를 산출할 수 있다. 이를 통해, 에이전트(120)는 최적의 주행 경로를 찾을 수 있는 최종 주행 정책을 결정하기 위한 데이터를 최종 주행 정책 결정부(160)에 제공할 수 있다.
- [59] 일 실시예에 따르면, 고전 경로 계획부(140)는 고전 경로 계획 알고리즘에 의하여 경로 라이브러리(110)에 저장된 복수의 주행 경로 중에서, 하나 이상의 주행 경로와 관련된 제2 수치를 산출할 수 있다. 예를 들어, 고전 경로 계획부(140)는 고전 경로 계획 알고리즘에 기초하여, 경로 라이브러리(110)에 포함된 하나 이상의 주행 경로의 각각에 대한 비용을 산출할 수 있다. 예를 들어, 하나 이상의 주행 경로에 대한 비용의 각각은 목표 지점까지의 거리에 대한 비용(goal distance cost), 주행 거리에 대한 비용(path distance cost), 타겟 속도에 대한 비용(target velocity cost) 및 충돌 비용(collision cost) 중 적어도 하나 또는 이를 합산한 비용을 포함할 수 있다.
- [60] 일 실시예에서, 고전 경로 계획 알고리즘은, 동적 창 접근 알고리즘(dynamic window approach algorithm), 트레젝토리 롤아웃 알고리즘(trjectory rollout algorithm), 포커스 트레젝토리 플래닝 알고리즘(focused trajectory planning algorithm) 또는 샘플링 기반 로컬 경로 플래너 알고리즘(sampling-based local trajectory planner algorithm) 중 적어도 하나를 포함할 수 있다. 이를 통해, 고전 경로 계획부(140)는 에이전트(120)와 마찬가지로 최적의 주행 경로를 찾을 수 있는 최종 주행 정책을 설립하기 위한 데이터를 최종 주행 정책 결정부(160)에 제공할 수 있다.
- [61] 일 실시예에 따르면, 최종 주행 정책 결정부(160)는 에이전트(120) 및 고전 경로 계획부(140)의 각각에서 산출된 수치에 기초하여 최적의 주행 경로를 탐색하는 에이전트의 최종 주행 정책을 결정할 수 있다. 예를 들어, 최종 주행 정책 결정부(160)는 경로 라이브러리(110)에 포함된 복수의 경로의 가치 및/또는 비용을 고려하여 에이전트의 최종 주행 정책을 결정할 수 있다. 여기서, 최종 주행 정책은, 대상 객체의 하나 이상의 상태의 각각에서의 주행 경로를 포함한 복수의 주행 경로를 포함할 수 있다.

- [62] 일 실시예에 따르면, 최종 주행 정책 결정부(160)는 산출된 수치 각각에 가중치를 적용하고 합산 점수를 산출할 수 있다. 예를 들어, 가치의 경우 가중치를 양수로, 비용의 경우 가중치를 음수로 적용하여 합산 점수를 산출할 수 있다. 이 경우, 가중치는 상황에 따라 산출된 수치 간 비중을 조절하기 위해 동적으로 조절될 수 있다.
- [63] 또한, 최종 주행 정책 결정부(160)는 산출한 합산 점수에 기초하여 주행 경로를 선정하는 최종 주행 정책을 설립하거나 결정할 수 있다. 예를 들어, 최종 주행 정책 결정부(160)는 합산한 점수가 최대가 되는 주행 경로를 선택하도록 최종 주행 정책을 결정할 수 있다.
- [64] 도 5는 본 개시의 일 실시예에 따른 강화학습 모델을 나타내는 예시도이다. 강화학습(Reinforcement Learning, RL)은 기계학습의 한 분야로, 에이전트(510)가 주어진 환경(500)에서의 현재 상태에 대해 선택 가능한 액션들 중 보상을 최대화하는 액션을 선택하는 학습 방법을 지칭할 수 있다. 여기서, 에이전트(510)는 강화학습의 대상이 되는 알고리즘 및/또는 기계학습 모델(예: 인공신경망 모델 등)을 포함할 수 있고, 주어진 환경을 탐색하여 현재의 상태에서 자신이 취할 수 있는 액션의 확률을 나타내는 정책(policy)을 수립할 수 있다. 여기서, 에이전트(510)는 본 명세서에서 개시된 에이전트(120)에 대응될 수 있다.
- [65] 본 개시에서, 강화학습은 연속적인 액션을 다룰 수 있는 알고리즘은 모두 포함할 수 있다. 일 실시예에서, 에이전트는 정책 기반 강화학습(policy-based RL)을 이용하여 학습될 수 있다. 여기서, 정책 기반 강화학습은 신경망을 이용하여 정책(policy)을 함수로 직접 모델링하는 학습 방식을 지칭할 수 있다. 이렇게 모델링된 정책 신경망에 상태(state)가 입력으로 들어오면, 이에 대한 액션(action)을 직접 출력할 수 있다.
- [66] 다른 실시예에서, 에이전트는 액터-크리틱(actor-critic) 알고리즘을 이용하여 학습될 수 있다. 여기서, 액터-크리틱 알고리즘은 정책 기반 강화학습을 통해 액션을 결정할 수 있고, 가치 함수(value function)를 이용하여 이 정책의 학습을 도와줄 수 있다. 예를 들어, A2C(Advantage Actor-Critic), A3C(Asynchronous Advantage Actor-Critic), PPO(Proximal Policy Optimization) 등이 액터-크리틱 알고리즘에 포함될 수 있다.
- [67] 즉, 본 개시는, 정책 기반 강화학습 알고리즘 및 액터-크리틱 알고리즘을 모두 포함할 수 있는데, 두 방식은 모두 정책 그래디언트(Policy Gradient, PG)를 이용할 수 있다. 여기서, 정책 그래디언트는 에이전트가 더 많은 보상(reward)을 받을 수 있도록 정책의 파라미터를 조금씩 바꿔가는 방식을 지칭할 수 있다. 이를 통해, 모델링된 정책 신경망은 최적의 정책으로 학습될 수 있다.
- [68] 이러한 정책 그래디언트 알고리즘은, 주어진 상태에 대하여 액션의 확률을 출력하는 방식인(Stochastic Policy Optimization) 및 주어진 상태에 대한 액션을 직접 출력하는 방식(Deterministic Policy Optimization)을 포함할 수 있다. 여기서, 액션의 확률을 출력하는 알고리즘(확률 정책 그래디언트)에는 A2C(Advantage Actor-

Critic), SAC(Soft Actor Critic), PPO(Proximal Policy Optimization), TRPO(Trust Region Policy Optimization) 등이 사용될 수 있으며, 액션을 직접 출력하는 알고리즘(결정 정책 그래디언트)에는 DDPG(Deep Deterministic Policy Gradient), TD3(Twin Delayed Deep Deterministic Policy Gradient) 등이 사용될 수 있으나, 이에 한정되지 않는다.

- [69] 본 개시에서, 상태(520)는 대상 객체의 현재 위치 정보를 지칭할 수 있다. 여기서, 현재 위치 정보는, GPS 정보, 특정 공간 내의 위치 정보 등과 같은 절대적인 위치 정보이거나 특정 참조 객체로부터의 상대적인 위치 정보를 지칭할 수 있다. 액션(540)은 주행 경로에 따라 주행하는 것을 포함할 수 있다. 또한, 보상(530)은 대상 객체의 액션에 대해 얻는 이익으로서, 대상 객체가 목표 지점에 도달하는 경우 갱신되는 에이전트의 기계학습 모델(예: 인공신경망 모델)의 가중치를 지칭할 수 있으나, 이에 한정되지 않으며, 강화학습의 목적 함수를 최대화할 수 있는 파라미터는 보상을 통해 갱신될 수 있다. 환경(500)은 대상 객체의 주행과 연관되거나 대상 객체(예: 로봇 등)의 주행을 나타내거나 특징화하는 임의의 환경을 지칭하며, 예를 들어, 대상 객체의 주행과 연관된 map 정보, 주행 경로 주변에 존재하는 방해물에 대한 정보 등을 포함할 수 있다. 또한, 환경(500)은, 에이전트의 제어 명령에 따라 주행하는 대상 객체에 대한 정보(예: 위치, 속도 등)도 포함할 수 있다. 또한, 환경(500)은 후술하는 환경(610)을 지칭할 수 있다.
- [70] 일 실시예에 따르면, 에이전트(510)가 주어진 환경(500)로부터 특정 상태를 수신하면(또는 에이전트가 모니터링을 통해 특정 상태를 획득하게 되면), 에이전트(510)는 상태에 따라 제어 명령을 생성하고, 생성된 제어 명령을 환경(500)의 대상 객체에 전달할 수 있다. 이에 응답하여, 환경(500)의 대상 객체는 액션에 대한 보상(530)을 에이전트(510)에 제공할 수 있다. 이러한 과정을 반복함으로써, 에이전트(510)는 환경(500)과 상호작용을 하며 보상을 많이 취할 수 있는 액션들이 결정하도록 학습될 수 있다.
- [71] 일 실시예에 따르면, 상술된 강화학습의 대상을 표현하기 위하여, 의사결정 과정을 모델링한 MDP(Markov Decision Process)가 이용될 수 있다. MDP는 시간이 진행함에 따라 상태가 확률적으로 변화하는 과정을 나타내는 Markov Process에서 보상(530), 액션(540) 및 정책(policy)이라는 개념이 추가된 의사결정 모델이다.
- [72] 이에 따라, MDP를 기초로 대상 객체의 네비게이션 작업 M은 수학식 1로 표현될 수 있다.
- [73] [수식1]

$$M=(O,A,r,P,\gamma)$$
- [74]
- [75] 위 수학식 1의 작업 M에서, O는 관측값(observation)을 나타내고, A는 액션(540)의 공간을 나타낼 수 있다. 여기서, 관측값은 상태(520)를 지칭할 수 있는 대상 객

체의 상태와 연관된 임의의 정보, 예를 들어, 현재 위치, 주행 속도, 주행 방향 등을 나타낼 수 있고, 액션(540)은 대상 객체가 어떤 경로로 주행할지에 대한 제어 명령을 나타낼 수 있다. 또한, 작업 M에서 r은 대상 객체의 액션에 대한 보상 함수를 나타낼 수 있다. 또한, 작업 M에서 P 및 γ 의 각각은 상태 전이 확률과 감가율을 나타낼 수 있다.

[76] 도 6은 본 개시의 일 실시예에 따른 경로 라이브러리(110), 에이전트(120) 및 환경(610) 사이의 상호 작용(interaction)에 대한 예시는 나타내는 도면이다. 일 실시예에 따르면, 에이전트(120)는 강화학습의 대상이 되는 알고리즘 및/또는 기계학습 모델을 포함할 수 있고, 대상 객체의 주어진 환경(610)과 상호 작용하여 현재 상태에서 자신이 취할 수 있는 액션(620)의 확률을 나타내는 정책을 수립하거나 결정할 수 있다. 여기서, 액션(620)은 경로 라이브러리(110)에 포함된 복수의 주행 경로 중 하나 이상의 주행 경로를 지칭할 수 있다. 예를 들어, 에이전트(120)는 주행 경로 선택 알고리즘을 포함할 수 있고, 대상 객체(예: 자율 이동 로봇)의 주어진 환경(610)을 탐색하여 목적지에 충돌하지 않고 도달하는 액션(620)의 확률을 나타내는 정책을 수립하거나 결정할 수 있다. 이 경우, 에이전트(120)의 주행 경로는 경로 라이브러리(110)에 포함된 경로일 수 있다.

[77] 일 실시예에 따르면, 경로 라이브러리(110)는 대상 객체의 주어진 환경(610)을 반영하여 업데이트될 수 있다(630). 예를 들어, 경로 라이브러리(110)는 환경(610)으로부터 대상 객체(예: 로봇)의 상태와 액션에 대한 데이터를 수신하고, 목적지까지 대상 객체가 주행하였을 때 얻게 될 기대 보상을 최대화하는 경로를 포함하도록 업데이트될 수 있다. 즉, 최대화된 기대 보상을 가진 주행 경로가 경로 라이브러리(110)에 포함되도록 경로 라이브러리(110)가 업데이트될 수 있다. 이 경우, 대상 객체의 상태는 대상 객체가 환경(610) 내에서 자율 주행하는데 필요한 임의의 정보를 포함할 수 있는데, 이에 한정되지 않으나, 예를 들어, 주행 범위(range), 위치(positioning), 현재 속도(current velocity) 등을 포함할 수 있다. 또한, 대상 객체의 액션은 대상 객체가 주행할 경로를 지칭할 수 있다.

[78] 도 7은 본 개시의 일 실시예에 따라 경로 라이브러리(110)가 업데이트되는(720) 예시를 나타내는 도면이다. 일 실시예에 따르면, 경로 라이브러리(110)는 액션 가치 함수(action-value function)를 이용하여 업데이트될 수 있다(720). 여기서, 액션 가치 함수란 대상 객체의 특정 상태에서 미래에 받을 기대 보상의 합을 지칭할 수 있다. 또한, 액션 가치 함수는 대상 객체의 특정 상태에서 특정 정책에 따라 특정 액션을 취했을 때 받는 반환 값의 기댓값을 지칭할 수 있다. 경로 라이브러리(110)를 업데이트하기 위한 액션 가치 함수는 하기의 수학식 2와 같다.

[79] [수식2]

$$Q_{\theta}(s_t, a_t) = \sum_{t=1}^{End} E_{s_t \sim D, a_t \sim \pi} [r(s_t, a_t)]$$

[80]

- [81] 위 수학식 2에서, Q 는 현재 단계(t)에서 마지막 단계까지 대상 객체(예: 로봇)가 주행하였을 때 획득할 수 있는 기대 보상을 지칭할 수 있다. 이러한 기대 보상을 획득하기 위해 기계학습 모델이 이용될 수 있다.
- [82] 위 수학식 2에서, θ 는 인공신경망 파라미터를 지칭할 수 있다. 예를 들어, θ 는 Q 네트워크 파라미터(Q network parameter)를 포함할 수 있다. s 는 대상 객체의 현재 상태를 지칭할 수 있다. 예를 들어, 대상 객체의 현재 상태는 대상 객체와 연관된 환경 내에서 주행하는데 필요한 임의의 정보를 포함할 수 있는데, 이에 한정되지 않으나, 예를 들어, 대상 객체의 주행 범위, 대상 객체의 현재 위치, 대상 객체의 현재 주행 속도를 포함할 수 있다.
- [83] 위 수학식 2에서, a 는 대상 객체의 액션을 지칭할 수 있다. 예를 들어, a 는 대상 객체가 주행할 주행 경로를 지칭할 수 있다. E 는 기댓값(expectation)을 지칭할 수 있으며, D 는 재생 버퍼(replay buffer)를 지칭할 수 있다. r 은 대상 객체가 특정 상태 s 에서 특정 액션 a 를 했을 때 얻게 될 보상을 지칭할 수 있다. 또한, π 는 대상 객체의 주행 정책을 지칭할 수 있고, t 는 대상 객체의 주행 과정의 단계를 지칭할 수 있다.
- [84] 일 실시예에 따르면, 경로 라이브러리(110)는 기계학습 모델을 이용하여 업데이트될 수 있다. 예를 들어, 경로 라이브러리(110)는 대상 객체의 상태와 액션을 입력으로 하여 현재 단계에서 마지막 단계까지의 기대 보상을 출력하도록 구성된 제1 기계학습 모델(710)에 의해 업데이트될 수 있다(720). 여기서, 제1 기계학습 모델(710)은 대상 객체의 학습 상태와 학습 액션을 포함한 학습 정보 및 해당 학습 정보에 대응하는 학습 기대 보상에 대한 정보를 이용하여 학습될 수 있다. 즉, 기계학습 모델(710)은 대상 객체의 학습 상태와 학습 액션을 입력 받고, 기대 보상을 출력할 수 있는데, 출력된 기대 보상과 주어진 학습 기대 보상 사이의 손실 값을 최소화하도록, 학습될 수 있다.
- [85] 일 실시예에 따르면, 제1 기계학습 모델(710)은 경로 라이브러리(110)를 업데이트하기 위하여 상술한 액션 가치 함수를 이용하여 구축되거나 학습될 수 있다. 예를 들어, 제1 기계학습 모델(710)은 상술한 액션 가치 함수에 의해 산출된 기대 보상을 최대화하는 방향으로 경로 라이브러리(110)를 업데이트 할 수 있다. 이때, 기대 보상을 최대화하기 위해 경사 상승법(gradient ascent)을 이용할 수 있다. 경사 상승법은 함수의 기울기를 구하고 기울기의 방향으로 계속 이동시켜 극값에 이를 때까지 반복시키는 것을 지칭할 수 있다. 경사 상승법을 이용하여 액션 가치 함수의 기대 보상을 최대화하기 위한 수식은 하기의 수학식 3과 같다.
- [86] [수식3]
- $$E_{s \sim D, a \sim \pi}[\nabla_a Q_\theta(s, a)]$$
- [87]
- [88] 위 수학식 3에서, 수학식 2에서 정의된 변수의 정보와 동일한 정보가 사용될 수 있다. 제1 기계학습 모델(710)의 학습 과정을 통해 출력된 경로, 즉, 숙련된 주

행을 위해 필요한 경로가, 경로 라이브러리(110)에 포함되도록 경로 라이브러리(110)가 업데이트될 수 있다. 나아가, 경로 라이브러리(110)는 대상 객체의 상태를 반영하여 실시간으로 업데이트될 수 있다.

[89] 도 8은 본 개시의 일 실시예에 따라 에이전트(120)의 초기 주행 정책이 학습되는(820) 예시를 나타내는 도면이다. 일 실시예에 따르면, 에이전트(120)의 초기 주행 정책을 학습시키기 위해서(820), 제2 기계학습 모델(810)이 이용될 수 있다. 보다 상세하게, 에이전트(120)의 초기 주행 정책 설립을 위한 제2 기계학습 모델(810)은 경로 라이브러리(110)에 포함된 주행 경로와 대상 객체의 상태 및 액션을 포함한 환경(610) 및 대상 객체의 상태 및 액션에 따른 기대보상을 이용하여 초기 주행 정책을 설립하거나 학습할 수 있다.

[90] 일 실시예에 따르면, 제2 기계학습 모델(810)은 에이전트(120)의 초기 주행 정책 설립을 위해 액션 가치 함수를 이용하여 구축되거나 학습될 수 있다. 제2 기계학습 모델(810)은 대상 객체의 상태와 액션을 입력으로 하여 기대 보상을 출력하도록 구성될 수 있는데, 예를 들어, 제2 기계학습 모델(810)은 제1 기계학습 모델(710)과 동일한 구성일 수 있으며, 도 7에서 상술한 액션 가치 함수를 이용할 수 있다.

[91] 일 실시예에 따르면, 제2 기계학습 모델(810)은 액션 가치 함수의 기대 보상을 최대화하기 위해 소프트맥스(softmax) 함수에 의해 산출된 확률 값에 기초하여 에이전트(120)의 초기 주행 정책을 학습시키도록(820) 구성될 수 있다. 여기서, 소프트맥스 함수는 로지스틱 함수의 다차원 일반화로, 기계학습 모델을 구축하거나 학습하는데 사용되는 확률 분포를 산출하기 위한 활성화 함수를 지칭할 수 있다. 예를 들어, 소프트맥스 함수의 일 예로서, 아래 수학식 4가 이용될 수 있다.

[92] [수식4]

$$\pi(A^i | s) = \frac{E_{\epsilon}[\exp(\frac{Q_{\theta}(s, A^i)}{\alpha})]}{\sum E_{\epsilon}[\exp(\frac{Q_{\theta}(s, A^i)}{\alpha})]}$$

[93]

[94] 위 수학식 4에서, π 는 객체(예: 로봇)의 주행 정책을 지칭할 수 있다. ϵ 는 가우시안 분포(gaussian distribution)를 위한 가우시안 노이즈(gaussian noise)를 지칭할 수 있다. α 는 엔트로피 정규화 계수(entropy regularization coefficient)를 지칭할 수 있다. 또한, A 는 객체(예: 로봇)의 행동을 지칭할 수 있다. 예를 들어, A 는 객체(예: 로봇)가 주행할 경로를 지칭할 수 있다.

[95] 이와 같이, 소프트맥스 함수를 이용하여 기대 보상에 대한 확률 값을 산출하도록 구성된 제2 기계학습 모델(810)은 에이전트(120)의 탐험(exploration)을 고려한 초기 주행 정책을 설립하고 학습시키는데 사용될 수 있다. 여기서, 제2 기계학습 모델(810)은 대상 객체의 학습 상태와 학습 액션을 포함한 학습 정보 및 해당 학

습 정보에 대응하는 기대보상에 대한 학습 확률값에 대한 정보를 이용하여 학습될 수 있다. 즉, 제2 기계학습 모델(810)은 대상 객체의 학습 상태와 학습 액션을 입력 받고, 기대 보상에 대한 확률값을 출력할 수 있는데, 출력된 확률값과 주어진 학습 확률값 사이의 손실 값을 최소화하도록, 학습될 수 있다.

- [96] 도 8에서, 제2 기계학습 모델(810)은 에이전트(120)의 외부 구성으로 도시되어 있으나, 이에 한정되지 않으며, 제2 기계학습 모델(810)은 에이전트(120)의 일부 구성에 포함될 수 있다.
- [97] 도 9은 본 개시의 다른 실시예에 따라 에이전트(120)의 초기 주행 정책이 학습되는(920) 예시를 나타내는 도면이다. 일 실시예에 따르면, 에이전트(120)의 초기 주행 정책을 학습시키기 위해서(920), 기계학습 모델이 이용될 수 있다. 보다 상세하게, 에이전트(120)의 초기 주행 정책 설립을 위한 제3 기계학습 모델(910)은 경로 라이브러리(110)에 포함된 경로와 대상 객체의 상태를 포함한 환경(610)으로부터 초기 주행 정책을 설립하거나 학습할 수 있다.
- [98] 일 실시예에 따르면, 제3 기계학습 모델(910)은 대상 객체의 상태 및 경로 라이브러리에 포함된 복수의 주행 경로를 입력으로 하고, 경로 라이브러리에 포함된 복수의 주행 경로의 각각을 선택할 확률을 출력하도록 구성될 수 있다. 여기서, 제3 기계학습 모델(910)은 대상 객체의 학습 상태 및 해당 학습 상태에 대응하는 복수의 학습 주행 정보, 복수의 학습 주행 정보의 각각을 선택할 학습 확률값에 대한 학습 정보를 이용하여 학습될 수 있다. 즉, 제3 기계학습 모델(910)은 대상 객체의 학습 상태 및 복수의 주행 정보를 입력 받고, 복수의 주행 정보의 각각에 대한 확률값을 출력할 수 있는데, 출력된 확률값과 주어진 학습 확률값 사이의 손실 값을 최소화하도록, 학습될 수 있다. 이와 같이 출력된 확률은 에이전트(120)가 주행 경로를 선택하기 위한 데이터로서 활용될 수 있다.
- [99] 일 실시예에 따르면, 제3 기계학습 모델(910)은 경로 라이브러리에 포함된 복수의 주행 경로 중에서, 최대 확률을 가진 주행 경로를 선택하도록 구성될 수 있다. 선택된 주행 경로를 이용하여 에이전트(120)의 초기 주행 정책이 설립되거나 학습될 수 있다. 이를 위해, 제3 기계학습 모델(910)은 최대 확률을 가진 주행 경로를 선택하도록 구성된 목적 함수(J)를 이용하여 구축되거나 학습될 수 있는데, 예를 들어, 목적 함수(J)를 최대화하도록 학습될 수 있다. 에이전트(120)의 초기 주행 정책 설립을 위한 목적 함수(J)는 하기의 수식식 5와 같다.

[100] [수식5]

$$J_{\pi}(\Phi)=E_{s_t \sim D}[E_{a_t \sim \pi_t}[\alpha \log(\pi_t(a_t | s_t)) - Q_{\theta}(s_t, a_t)]]$$

[101] $Q(s_t, a_t)=r(s_t, a_t)+\gamma E_{s_{t+1} \sim p(s_{t+1}|s_t, a_t)}[V(s_{t+1})]$

[102] $V(s_t)=E_{a_t \sim \pi_t}[Q(s_t, a_t)-\alpha \log(\pi(a_t | s_t))]$

[103]

- [104] 위 수학식 5에서, J 는 목적 함수, Q 는 행동 가치 함수(Q 함수), V 는 상태 가치 함수를 지칭할 수 있다. 또한, π 는 객체(예: 로봇)의 주행 정책을 지칭할 수 있다. ϕ 는 인공신경망 파라미터를 지칭할 수 있다. 예를 들어, ϕ 는 행위자 네트워크 파라미터(actor network parameter)를 포함할 수 있다. θ 는 다른 인공신경망 파라미터를 지칭할 수 있다. 예를 들어, θ 는 Q 네트워크 파라미터를 포함할 수 있다.
- [105] 위 수학식 5에서, E 는 기댓값을 지칭할 수 있고, D 는 재생 버퍼를 지칭할 수 있다. 또한 p 는 미지의 상태 전이 확률(unknown state transition probability)을 지칭할 수 있고, α 는 엔트로피 정규화 계수를 지칭할 수 있다.
- [106] 이와 같이 제3 기계학습 모델(910)은 목적 함수(J)를 최대화하도록 에이전트(120)의 초기 주행 정책을 학습시킴으로써, 경로 라이브러리(110)에 포함된 복수의 주행 경로 중 최대 기대 보상을 가진 주행 경로가 선택되도록 에이전트(120)의 초기 주행 정책이 설립되거나 학습될 수 있다.
- [107] 도 9에서, 제3 기계학습 모델(910)은 에이전트(120)의 외부 구성으로 도시되어 있으나, 이에 한정되지 않으며, 제3 기계학습 모델(910)은 에이전트(120)의 일부 구성에 포함될 수 있다.
- [108] 도 10은 본 개시의 일 실시예에 따라 주행 경로에 대한 수치를 산출하고, 최종 주행 정책(1080)을 결정하는(1070) 예시를 나타내는 도면이다. 일 실시예에 따르면, 초기 주행 정책이 학습된 에이전트(120)는 경로 라이브러리(110)에 포함된 복수의 주행 경로를 입력받아(1010), 주행 경로 각각에 대한 가치(1030)를 산출할 수 있다(1020). 이 경우, 주행 경로 각각에 대한 가치(1030)는 주행 경로 각각에 대한 확률을 포함할 수 있다.
- [109] 일 실시예에 따르면, 고전 경로 계획부(140)는 고전 경로 계획 알고리즘에 기초하여, 경로 라이브러리(110)에 포함된 복수의 주행 경로를 입력받아(1040), 복수의 주행 경로의 각각에 대한 비용(1060)을 산출할 수 있다(1050). 이 경우, 주행 경로 각각에 대한 비용(1060)은 목표 지점까지의 거리에 대한 비용, 주행 거리에 대한 비용, 타겟 속도에 대한 비용 또는 충돌 비용 중 적어도 하나이거나 이러한 비용들을 합산한 비용일 수 있다.
- [110] 일 실시예에 따르면, 고전 경로 계획부(140)는 비용 산출을 위한 고전 경로 계획 알고리즘으로서, 동적 창 접근 알고리즘(dynamic window approach algorithm, DWA)을 이용할 수 있다. 동적 창 접근 알고리즘은 지역 경로계획에 해당하며 해당 객체의 동적 상태를 반영하여 충돌 회피가 가능한 목적지 추종 알고리즘으로, 전역 경로 제어 알고리즘과 융합된 하이브리드 제어 형태로 적용할 수 있다. 동적 창 접근 알고리즘은 목적지에 대한 방향, 속도, 장애물과의 거리를 평가하여 최적의 선속도와 각속도를 도출함으로써, 장애물을 회피하고 목적지에 도달하기 위해 이용될 수 있다.
- [111] 일 실시예에 따르면, 고전 경로 계획부(140)는 비용 산출을 위한 고전 경로 계획 알고리즘으로서, 트래젝토리 롤아웃 알고리즘(trjectory rollout algorithm)을 이용할 수 있다. 트래젝토리 롤아웃 알고리즘은 실시간으로 그래디언트 기술을 사용

하여 비용 맵에서 전역적으로 최적의 경로를 계획하고, 예측 동적 컨트롤러를 이용하여 로컬 속도 명령을 계산하기 위해 이용될 수 있다.

- [112] 일 실시예에 따르면, 고전 경로 계획부(140)는 비용 산출을 위한 고전 경로 계획 알고리즘으로서, 포커스 트래젝토리 플래닝 알고리즘(focused trajectory planning algorithm)을 이용할 수 있다. 포커스 트래젝토리 플래닝 알고리즘은 2단계 모션 플래닝 시스템으로서 레퍼런스 주행 경로 플래닝(reference trajectory planning)과 트래킹 주행 경로 플래닝(tracking trajectory planning)으로 구성될 수 있다. 레퍼런스 주행 경로 플래닝은 밀집 격자 샘플링 및 최적화 기술을 사용하여 인간과 유사한 주행 경로를 생성할 수 있다. 트래킹 주행 경로 플래닝은 동적 제약 조건을 더욱 충족하는 주행 경로를 생성, 평가 및 선택하는데 이용될 수 있다. 포커스 트래젝토리 플래닝 알고리즘은 레퍼런스 경로(reference path)를 기준으로 하여 주행 경로를 생성하므로, 경로 라이브러리(110)는 대상 객체의 현재 위치에서 전방 주시점으로 향하는 목표 선속도, 목표 회전속도를 레퍼런스로 하고 그 속도와의 차이를 포함하도록 구성될 수 있다.
- [113] 일 실시예에 따르면, 고전 경로 계획부(140)는 비용 산출을 위한 고전 경로 계획 알고리즘으로서 샘플링 기반 로컬 경로 플래너 알고리즘(sampling-based local trajectory planner algorithm)을 이용할 수 있다. 샘플링 기반 로컬 경로 플래너 알고리즘은 2단계 모션 플래닝 아키텍처를 사용하는 알고리즘으로, 첫 번째 단계에서 기준 경로를 세분화하기 위해 서포트 벡터 머신(support vector machine, SVM) 기반의 접근 방식이 이용될 수 있으며, 두 번째 단계에서는 정제된 기준 주행 경로와 일치하는 터미널 상태 집합이 샘플링되어 주행 경로를 예측하기 위해 이용될 수 있다. 이와 마찬가지로, 샘플링 기반 로컬 경로 플래너 알고리즘의 경우에도 레퍼런스 경로를 기준으로 하여 주행 경로가 생성되므로, 경로 라이브러리(110)는 대상 객체의 현재 위치에서 전방 주시점으로 향하는 목표 선속도, 목표 회전속도를 레퍼런스로 하고 그 속도와의 차이를 포함하도록 구성될 수 있다.
- [114] 일 실시예에 따르면, 최종 주행 정책 결정부(160)는 산출된 가치(1030) 및 비용(1060)의 각각에 가중치를 적용하여 합산한 수치가 가장 큰 주행 경로를 선택하도록 에이전트의 최종 주행 정책(1080)을 결정할 수 있다(1070). 이 경우, 적용되는 가중치는 컨텍스트에 따라 산출된 가치(1030)와 비용(1060) 사이의 비중을 고려하여 동적으로 조절될 수 있다.
- [115] 도 11은 본 개시의 일 실시예에 따른 하이브리드 주행 알고리즘 기반 주행 정책 결정 방법(1100)의 예시를 나타내는 흐름도이다. 방법(1100)은 정보 처리 시스템의 적어도 하나의 프로세서(예: 프로세서(320))에 의해 수행될 수 있다. 도시된 바와 같이, 방법(1100)은 경로 라이브러리를 이용하여 학습된 에이전트의 초기 주행 정책을 획득함으로써 개시될 수 있다(S1110).
- [116] 일 실시예에서, 경로 라이브러리는 전문가 에이전트의 주행 경로를 포함하도록 업데이트 될 수 있다. 또한, 경로 라이브러리는 대상 객체의 상태와 액션을 입력으로 하여 기대 보상을 출력하도록 구성된 제1 기계학습 모델에 의해 업데이트

될 수 있다. 이 경우, 제1 기계학습 모델은 경사 상승법을 이용하여 기대 보상을 최대화하는 방향으로 경로 라이브러리를 업데이트하도록 구성될 수 있다.

- [117] 일 실시예에서, 에이전트의 초기 주행 정책은 경로 라이브러리에 포함된 복수의 주행 경로 중에서, 가장 높은 보상을 주는 주행 경로를 선택하도록 학습될 수 있다. 에이전트의 초기 주행 정책은 대상 객체의 상태와 액션을 입력으로 하여 기대 보상을 출력하도록 구성된 제2 기계학습 모델을 이용하여 학습될 수 있다. 이 경우, 제2 기계학습 모델은 소프트맥스 함수에 의해 산출된 확률 값에 기초하여, 에이전트의 초기 주행 정책을 학습시키도록 구성될 수 있다.
- [118] 다른 실시예에서, 에이전트의 초기 주행 정책은 대상 객체의 상태 및 경로 라이브러리에 포함된 복수의 주행 경로를 입력으로 하고, 복수의 주행 경로의 각각을 선택할 확률을 출력하도록 구성된 제3 기계학습 모델을 이용하여 학습될 수 있다. 이 경우, 제3 기계학습 모델은 복수의 주행 경로 중에서, 최대 확률을 가진 주행 경로를 선택하도록 에이전트의 초기 주행 정책을 학습시키도록 구성될 수 있다.
- [119] 그런 다음, 프로세서는 학습된 에이전트의 초기 주행 정책을 이용한 제1 방식에 의하여, 하나 이상의 주행 경로와 관련된 제1 수치를 산출할 수 있다(S1120). 이 경우, 하나 이상의 주행 경로는 경로 라이브러리에 포함된 경로일 수 있다. 일 실시예에서, 제1 수치는 하나 이상의 주행 경로에 대한 가치를 포함할 수 있고, 제1 방식은 학습된 에이전트의 초기 주행 정책에 기초하여 하나 이상의 주행 경로에 대한 가치를 산출하는 방식을 포함할 수 있다.
- [120] 그런 다음, 프로세서는 제2 방식에 의하여 하나 이상의 주행 경로와 관련된 제2 수치를 산출할 수 있다(S1130). 일 실시예에서, 제2 수치는 하나 이상의 주행 경로에 대한 비용을 포함할 수 있고, 제2 방식은 고전 경로 계획 알고리즘에 기초하여, 하나 이상의 주행 경로에 대한 비용을 산출하는 방식을 포함할 수 있다. 이 경우, 하나 이상의 주행 경로의 각각에 대한 비용은 목표 지점까지의 거리에 대한 비용, 주행 거리에 대한 비용, 타겟 속도에 대한 비용 및 충돌 비용을 합산한 비용을 포함할 수 있다.
- [121] 일 실시예에서, 고전 경로 계획 알고리즘은, 동적 창 접근 알고리즘, 트레젝토리아아웃 알고리즘, 레퍼런스 경로를 기준으로 주행 경로를 생성하는 알고리즘 등을 포함할 수 있다. 이 경우, 경로 라이브러리는 대상 객체의 현재 위치에서 전방 주시점으로 향하는 목표 선속도 및 목표 회전속도를 레퍼런스로 하고 그 속도와 차이를 포함하여 구성될 수 있다.
- [122] 마지막으로, 프로세서는 산출된 제1 수치 및 산출된 제2 수치에 기초하여 에이전트의 최종 주행 정책을 결정할 수 있다(S1140). 일 실시예에서, 제1 수치 및 제2 수치 각각에 가중치를 적용하여 합산한 수치가 가장 큰 주행 경로를 선택하도록 에이전트의 최종 주행 정책을 결정할 수 있다. 이 경우, 제1 수치와 제2 수치의 각각에 적용되는 가중치는 동적으로 조절될 수 있다.

- [123] 도 11에서 도시한 흐름도 및 상술한 설명은 일 예시일 뿐이며, 일부 실시예에서는 다르게 구현될 수 있다. 예를 들어, 일부 실시예에서는 각 단계의 순서가 바뀌거나, 일부 단계가 반복 수행되거나, 일부 단계가 생략되거나, 일부 단계가 추가될 수 있다. 예를 들어, 프로세서는 하나 이상의 주행 경로와 관련된 제1 수치 및 제2 수치를 동시에 산출할 수 있다. 즉, 단계 S1120 및 단계 S1130의 과정이 동시에 수행될 수 있다.
- [124] 상술한 방법은 컴퓨터에서 실행하기 위해 컴퓨터 판독 가능한 기록 매체에 저장된 컴퓨터 프로그램으로 제공될 수 있다. 매체는 컴퓨터로 실행 가능한 프로그램을 계속 저장하거나, 실행 또는 다운로드를 위해 임시 저장하는 것일 수도 있다. 또한, 매체는 단일 또는 수개 하드웨어가 결합된 형태의 다양한 기록수단 또는 저장수단일 수 있는데, 어떤 컴퓨터 시스템에 직접 접속되는 매체에 한정되지 않고, 네트워크 상에 분산 존재하는 것일 수도 있다. 매체의 예시로는, 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체, CD-ROM 및 DVD 와 같은 광기록 매체, 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical medium), 및 ROM, RAM, 플래시 메모리 등을 포함하여 프로그램 명령어가 저장되도록 구성된 것이 있을 수 있다. 또한, 다른 매체의 예시로, 애플리케이션을 유통하는 앱 스토어나 기타 다양한 소프트웨어를 공급 내지 유통하는 사이트, 서버 등에서 관리하는 기록매체 내지 저장매체도 들 수 있다.
- [125] 본 개시의 방법, 동작 또는 기법들은 다양한 수단에 의해 구현될 수도 있다. 예를 들어, 이러한 기법들은 하드웨어, 펌웨어, 소프트웨어, 또는 이들의 조합으로 구현될 수도 있다. 본원의 개시와 연계하여 설명된 다양한 예시적인 논리적 블록들, 모듈들, 회로들, 및 알고리즘 단계들은 전자 하드웨어, 컴퓨터 소프트웨어, 또는 양자의 조합들로 구현될 수도 있음을 통상의 기술자들은 이해할 것이다. 하드웨어 및 소프트웨어의 이러한 상호 대체를 명확하게 설명하기 위해, 다양한 예시적인 구성요소들, 블록들, 모듈들, 회로들, 및 단계들이 그들의 기능적 관점에서 일반적으로 위에서 설명되었다. 그러한 기능이 하드웨어로서 구현되는지 또는 소프트웨어로서 구현되는지의 여부는, 특정 애플리케이션 및 전체 시스템에 부과되는 설계 요구사항들에 따라 달라진다. 통상의 기술자들은 각각의 특정 애플리케이션을 위해 다양한 방식으로 설명된 기능을 구현할 수도 있으나, 그러한 구현들은 본 개시의 범위로부터 벗어나게 하는 것으로 해석되어서는 안된다.
- [126] 하드웨어 구현에서, 기법들을 수행하는 데 이용되는 프로세싱 유닛들은, 하나 이상의 ASIC들, DSP들, 디지털 신호 프로세싱 디바이스들(digital signal processing devices; DSPD들), 프로그램가능 논리 디바이스들(programmable logic devices; PLD들), 필드 프로그램가능 게이트 어레이들(field programmable gate arrays; FPGA들), 프로세서들, 제어기들, 마이크로제어기들, 마이크로프로세서들, 전자 디바이스들, 본 개시에 설명된 기능들을 수행하도록 설계된 다른 전자 유닛들, 컴퓨터, 또는 이들의 조합 내에서 구현될 수도 있다.

- [127] 따라서, 본 개시와 연계하여 설명된 다양한 예시적인 논리 블록들, 모듈들, 및 회로들은 범용 프로세서, DSP, ASIC, FPGA나 다른 프로그램 가능 논리 디바이스, 이산 게이트나 트랜지스터 로직, 이산 하드웨어 컴포넌트들, 또는 본원에 설명된 기능들을 수행하도록 설계된 것들의 임의의 조합으로 구현되거나 수행될 수도 있다. 범용 프로세서는 마이크로프로세서일 수도 있지만, 대안으로, 프로세서는 임의의 종래의 프로세서, 제어기, 마이크로제어기, 또는 상태 머신일 수도 있다. 프로세서는 또한, 컴퓨팅 디바이스들의 조합, 예를 들면, DSP와 마이크로프로세서, 복수의 마이크로프로세서들, DSP 코어와 연계한 하나 이상의 마이크로프로세서들, 또는 임의의 다른 구성의 조합으로서 구현될 수도 있다.
- [128] 펌웨어 및/또는 소프트웨어 구현에 있어서, 기법들은 랜덤 액세스 메모리(random access memory; RAM), 판독 전용 메모리(read-only memory; ROM), 비휘발성 RAM(non-volatile random access memory; NVRAM), PROM(programmable read-only memory), EPROM(erasable programmable read-only memory), EEPROM(electrically erasable PROM), 플래시 메모리, 콤팩트 디스크(compact disc; CD), 자기 또는 광학 데이터 스토리지 디바이스 등과 같은 컴퓨터 판독가능 매체 상에 저장된 명령들로서 구현될 수도 있다. 명령들은 하나 이상의 프로세서들에 의해 실행 가능할 수도 있고, 프로세서(들)로 하여금 본 개시에 설명된 기능의 특정 양태들을 수행하게 할 수도 있다.
- [129] 이상 설명된 실시예들이 하나 이상의 독립형 컴퓨터 시스템에서 현재 개시된 주제의 양태들을 활용하는 것으로 기술되었으나, 본 개시는 이에 한정되지 않고, 네트워크나 분산 컴퓨팅 환경과 같은 임의의 컴퓨팅 환경과 연계하여 구현될 수도 있다. 또 나아가, 본 개시에서 주제의 양상들은 복수의 프로세싱 칩들이나 장치들에서 구현될 수도 있고, 스토리지는 복수의 장치들에 걸쳐 유사하게 영향을 받게 될 수도 있다. 이러한 장치들은 PC들, 네트워크 서버들, 및 휴대용 장치들을 포함할 수도 있다.
- [130] 본 명세서에서는 본 개시가 일부 실시예들과 관련하여 설명되었지만, 본 개시의 발명이 속하는 기술분야의 통상의 기술자가 이해할 수 있는 본 개시의 범위를 벗어나지 않는 범위에서 다양한 변형 및 변경이 이루어질 수 있다. 또한, 그러한 변형 및 변경은 본 명세서에 첨부된 특허청구의 범위 내에 속하는 것으로 생각되어야 한다.

청구범위

- [청구항 1] 적어도 하나의 프로세서에 의해 수행되는, 하이브리드(hybrid) 주행 알고리즘 기반 주행 정책(policy) 결정 방법에 있어서,
경로 라이브러리(trajjectory library)를 이용하여 학습된 에이전트의 초기 주행 정책을 획득하는 단계;
상기 학습된 에이전트의 초기 주행 정책을 이용한 제1 방식에 의하여, 하나 이상의 주행 경로와 관련된 제1 수치를 산출하는 단계;
제2 방식에 의하여, 상기 하나 이상의 주행 경로와 관련된 제2 수치를 산출하는 단계; 및
상기 산출된 제1 수치 및 상기 산출된 제2 수치에 기초하여, 상기 에이전트의 최종 주행 정책을 결정하는 단계
를 포함하는, 주행 정책 결정 방법.
- [청구항 2] 제1항에 있어서,
상기 경로 라이브러리는,
전문가 에이전트(expert agent)의 주행 경로를 포함하도록 업데이트되는,
주행 정책 결정 방법.
- [청구항 3] 제1항에 있어서,
상기 경로 라이브러리는,
대상 객체의 상태(state)와 액션(action)을 입력으로 하여 기대 보상을 출력하도록 구성된 제1 기계학습 모델에 의해 업데이트되는,
주행 정책 결정 방법.
- [청구항 4] 제3항에 있어서,
상기 제1 기계학습 모델은,
경사 상승법(gradient ascent)을 이용하여 기대 보상을 최대화하는 방향으로 상기 경로 라이브러리를 업데이트하도록 구성된,
주행 정책 결정 방법.
- [청구항 5] 제1항에 있어서,
상기 에이전트의 초기 주행 정책은,
상기 경로 라이브러리에 포함된 복수의 주행 경로 중에서, 가장 높은 보상(reward)을 주는 주행 경로를 선택하도록 학습되는,
주행 정책 결정 방법.
- [청구항 6] 제1항에 있어서,
상기 에이전트의 초기 주행 정책은,
대상 객체의 상태(state)와 액션(action)을 입력으로 하여 기대 보상을 출력하도록 구성된 제2 기계학습 모델을 이용하여 학습되는,
주행 정책 결정 방법.
- [청구항 7] 제6항에 있어서,

상기 제2 기계학습 모델은,
소프트맥스(softmax) 함수에 의해 산출된 확률 값에 기초하여, 상기 에이전트의 초기 주행 정책을 학습시키도록 구성된,
주행 정책 결정 방법.

[청구항 8] 제1항에 있어서,
상기 에이전트의 초기 주행 정책은,
대상 객체의 상태 및 상기 경로 라이브러리에 포함된 복수의 주행 경로를 입력으로 하고, 상기 경로 라이브러리에 포함된 복수의 주행 경로의 각각을 선택할 확률을 출력하도록 구성된 제3 기계학습 모델을 이용하여 학습되는,
를 포함하는, 주행 정책 결정 방법.

[청구항 9] 제8항에 있어서,
상기 제3 기계학습 모델은,
상기 복수의 주행 경로 중에서, 최대 확률을 가진 주행 경로를 선택하도록 상기 에이전트의 초기 주행 정책을 학습시키도록 구성된,
주행 정책 결정 방법.

[청구항 10] 제1항에 있어서,
상기 하나 이상의 주행 경로는,
상기 경로 라이브러리에 포함된 경로인,
주행 정책 결정 방법.

[청구항 11] 제1항에 있어서,
상기 제1 수치는,
상기 하나 이상의 주행 경로에 대한 가치(value)를 포함하고,
상기 제1 방식은,
상기 학습된 에이전트의 초기 주행 정책에 기초하여, 상기 하나 이상의 주행 경로에 대한 가치를 산출하는 방식을 포함하는, 주행 정책 결정 방법.

[청구항 12] 제1항에 있어서,
상기 제2 수치는,
상기 하나 이상의 주행 경로에 대한 비용(cost)을 포함하고,
상기 제2 방식은,
고전 경로 계획 알고리즘에 기초하여, 상기 하나 이상의 주행 경로에 대한 비용을 산출하는 방식을 포함하는, 주행 정책 결정 방법.

[청구항 13] 제12항에 있어서,
상기 하나 이상의 주행 경로의 각각에 대한 비용은,

목표 지점까지의 거리에 대한 비용(goal distance cost), 주행 거리에 대한 비용(path distance cost), 타겟 속도에 대한 비용(target velocity cost) 및 충돌 비용(collision cost)을 합산한 비용을 포함하는, 주행 정책 결정 방법.

[청구항 14] 제1항에 있어서,
상기 에이전트의 최종 주행 정책을 결정하는 단계는,
상기 제1 수치 및 상기 제2 수치의 각각에 가중치를 적용하여 합산한 수치가 가장 큰 주행 경로를 선택하도록 상기 에이전트의 최종 주행 정책을 결정하는 단계
를 포함하는, 주행 정책 결정 방법.

[청구항 15] 제14항에 있어서,
상기 제1 수치와 상기 제2 수치의 각각에 적용되는 가중치는 동적으로 조절되는,
주행 정책(policy) 결정 방법.

[청구항 16] 제1항에 따른 방법을 컴퓨터에서 실행하기 위한 명령어들을 기록한 컴퓨터 판독 가능한 비일시적 기록매체.

[청구항 17] 정보 처리 시스템으로서,
통신 모듈;
메모리; 및
상기 메모리와 연결되고, 상기 메모리에 포함된 컴퓨터 판독 가능한 적어도 하나의 프로그램을 실행하도록 구성된 적어도 하나의 프로세서를 포함하고,
상기 적어도 하나의 프로그램은,
경로 라이브러리를 이용하여 학습된 에이전트의 초기 주행 정책을 획득하고,
상기 학습된 에이전트의 초기 주행 정책을 이용한 제1 방식에 의하여, 하나 이상의 주행 경로와 관련된 제1 수치를 산출하고,
제2 방식에 의하여, 상기 하나 이상의 주행 경로와 관련된 제2 수치를 산출하고,
상기 산출된 제1 수치 및 상기 산출된 제2 수치에 기초하여, 상기 에이전트의 최종 주행 정책을 결정하기 위한 명령어들을 포함하는,
정보 처리 시스템.

[청구항 18] 제17항에 있어서,
상기 에이전트의 초기 주행 정책은,
상기 경로 라이브러리에 포함된 복수의 주행 경로 중에서, 가장 높은 보상(reward)을 주는 주행 경로를 선택하도록 학습되는,
정보 처리 시스템.

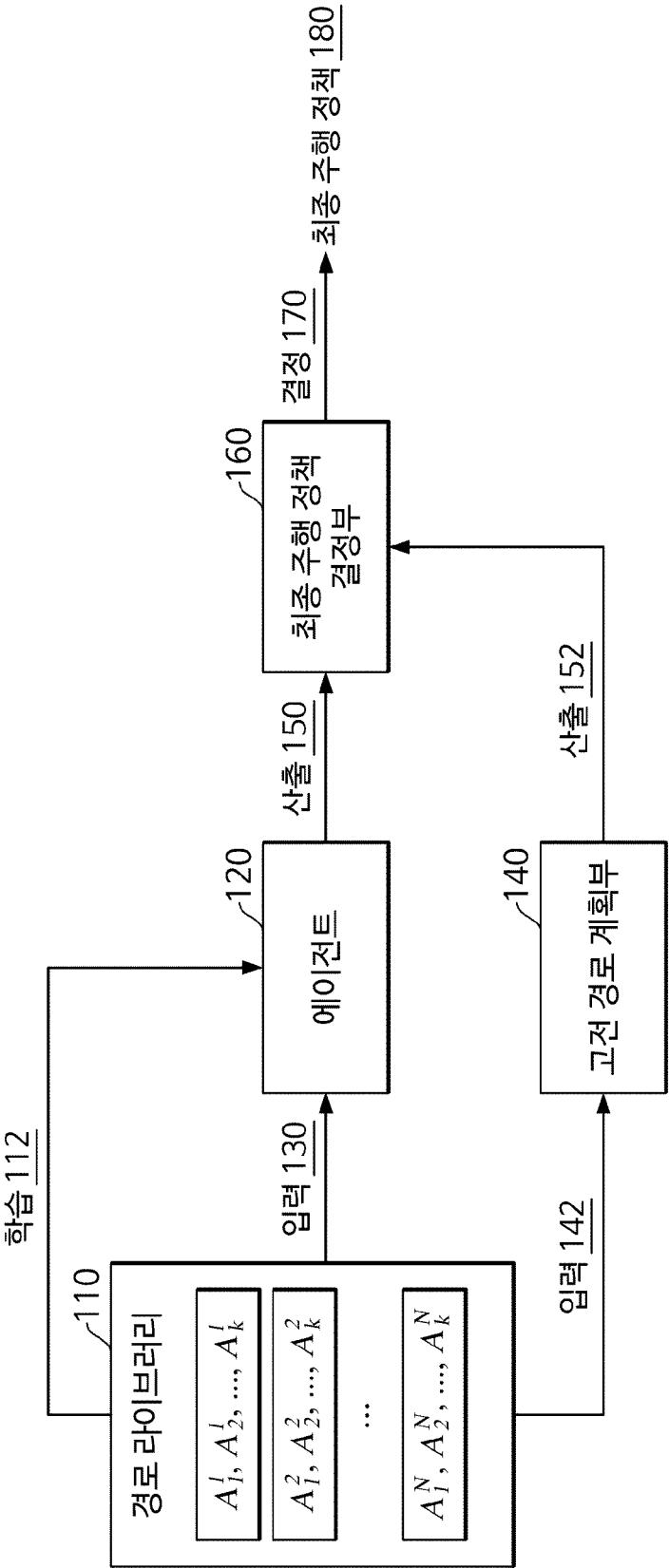
[청구항 19] 제17항에 있어서,

상기 제1 수치는,
상기 하나 이상의 주행 경로에 대한 가치(value)를 포함하고,
상기 제1 방식은,
상기 학습된 에이전트의 초기 주행 정책에 기초하여, 상기 하나 이상의 주행 경로에 대한 가치를 산출하는 방식을 포함하는,
정보 처리 시스템.

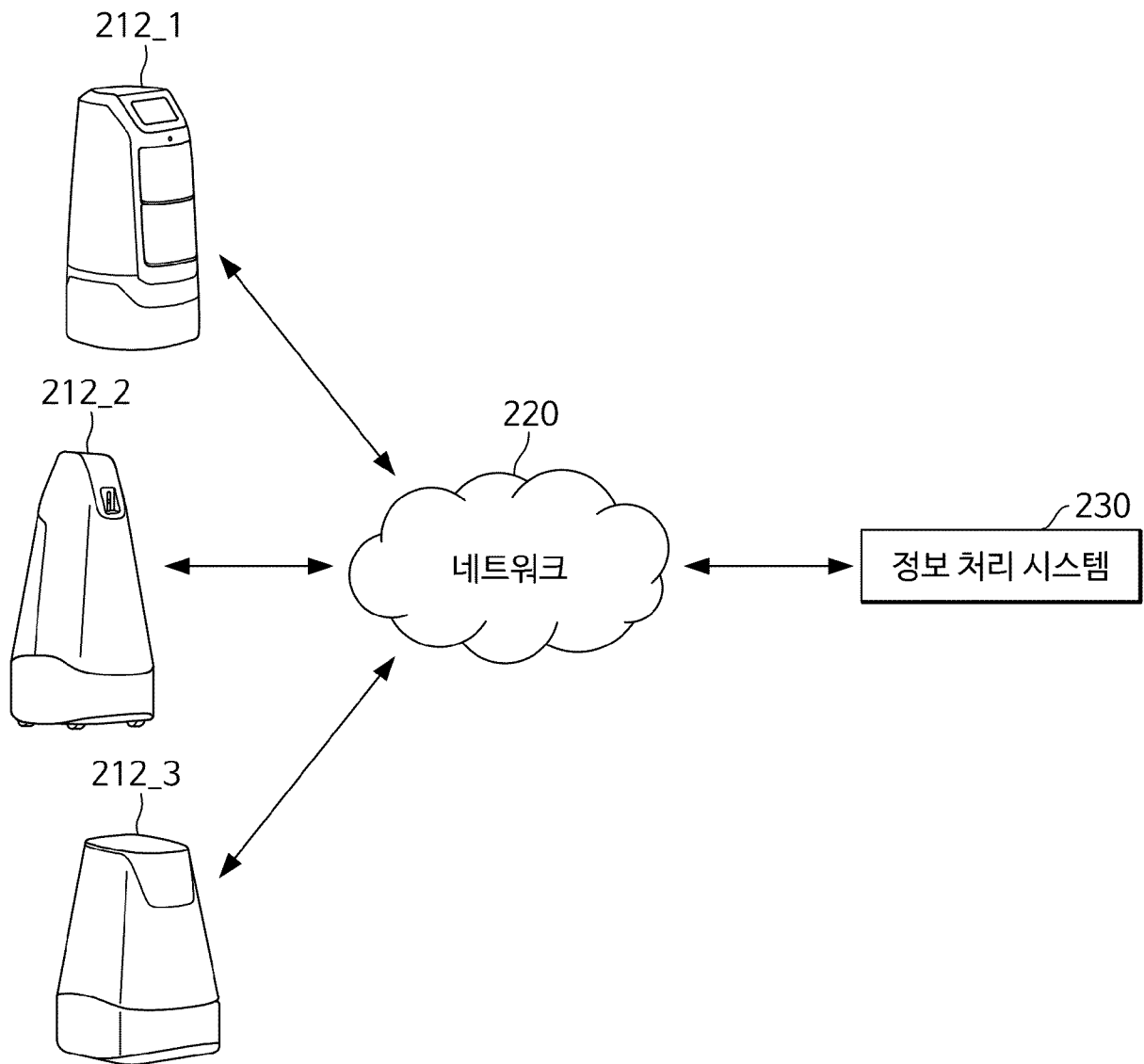
[청구항 20]

제17항에 있어서,
상기 제2 수치는,
상기 하나 이상의 주행 경로에 대한 비용(cost)을 포함하고,
상기 제2 방식은,
고전 경로 계획 알고리즘에 기초하여, 상기 하나 이상의 주행 경로에 대한 비용을 산출하는 방식을 포함하는,
정보 처리 시스템.

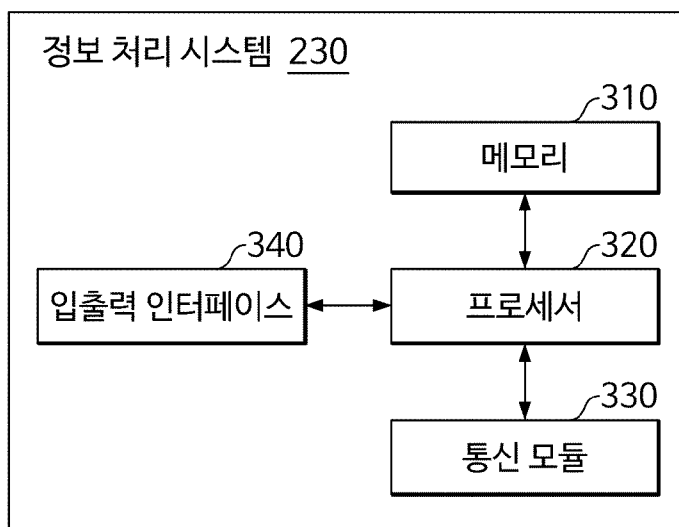
[도 1]



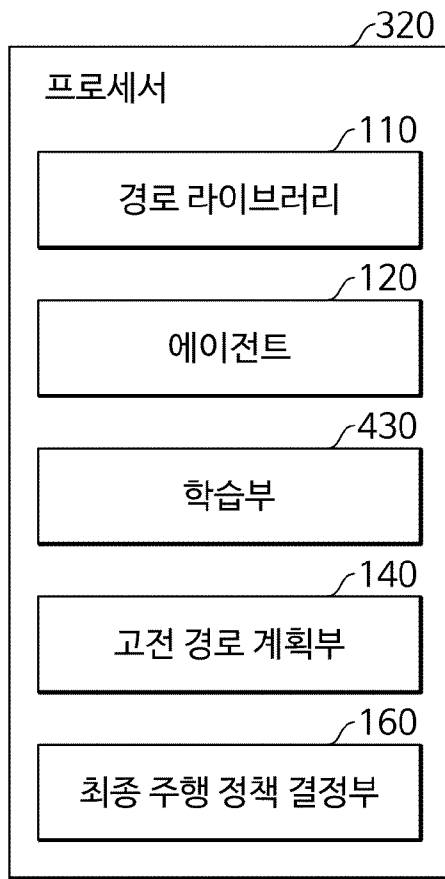
[도2]



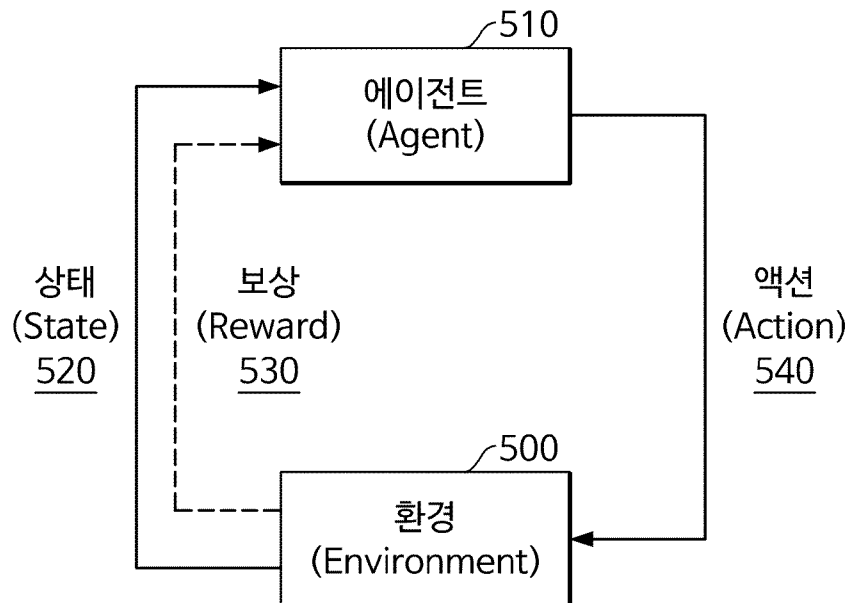
[도3]



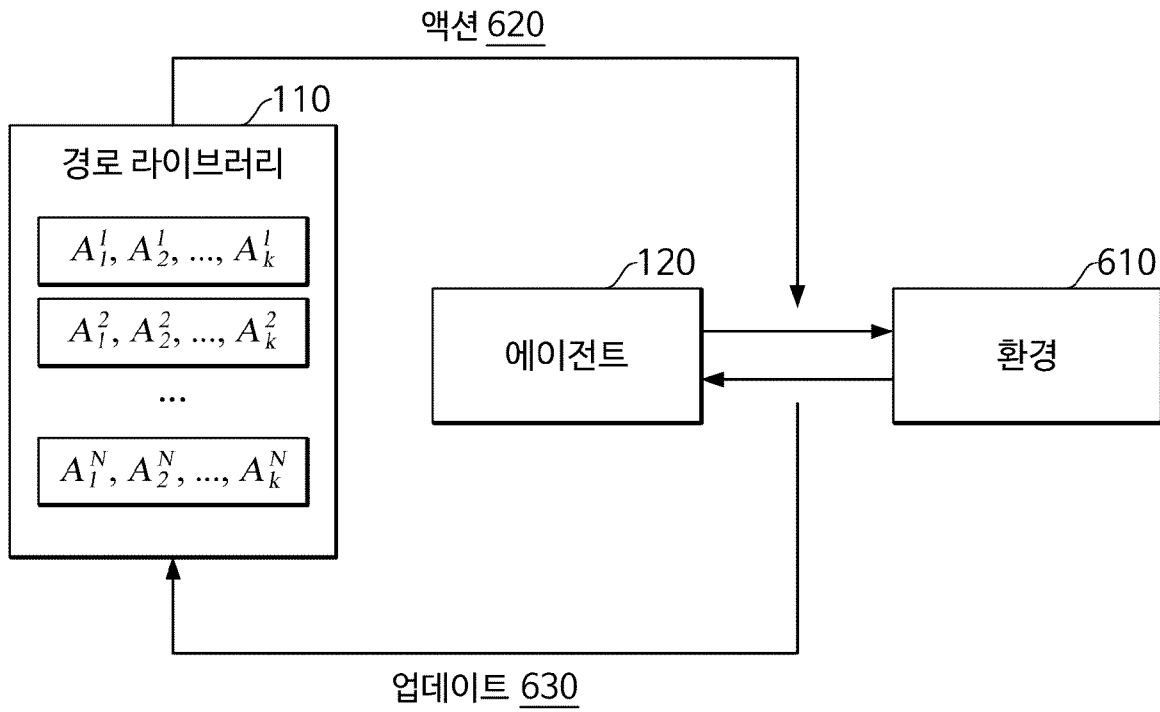
[도4]



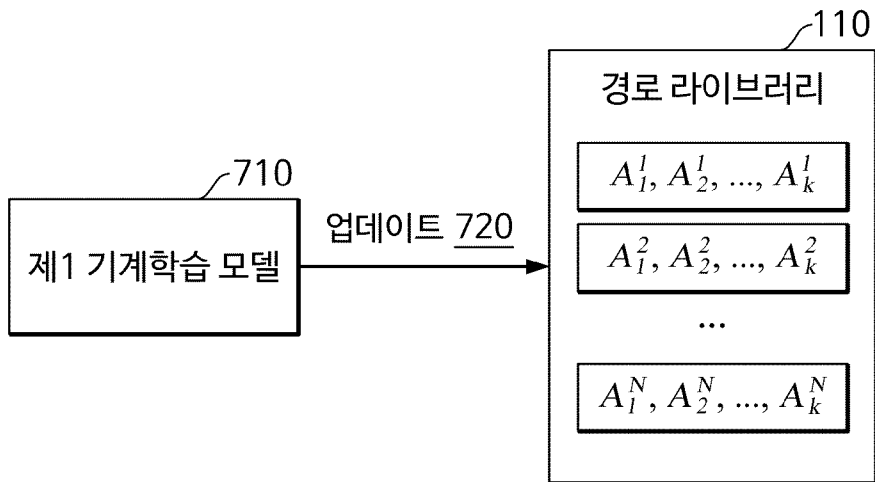
[도5]



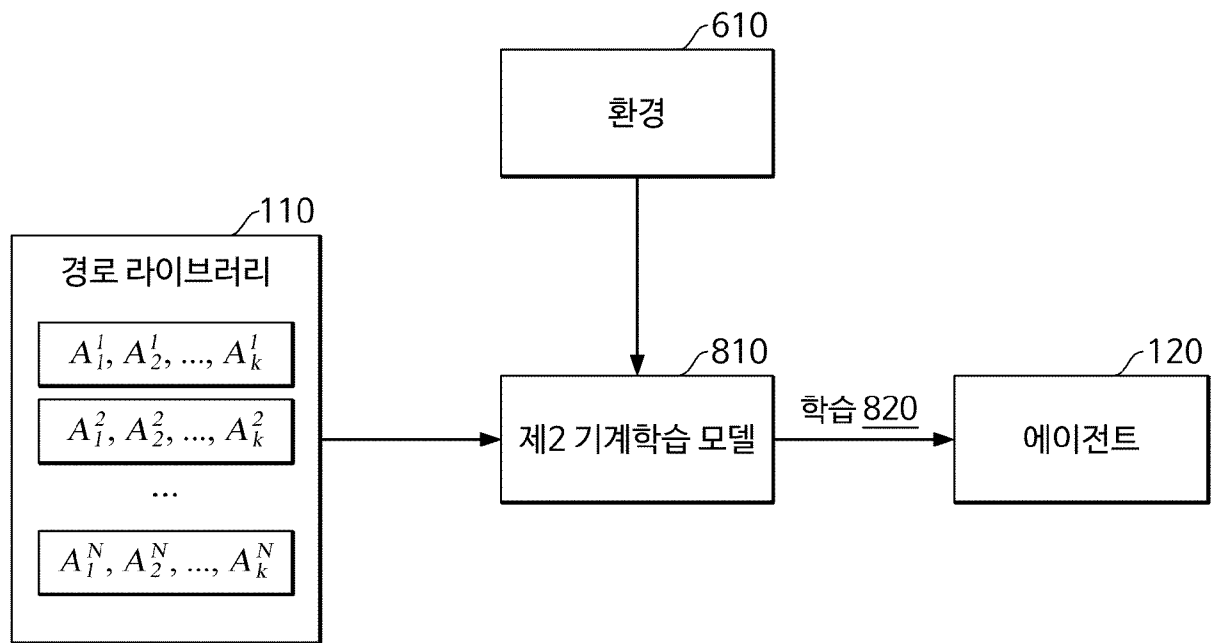
[도6]



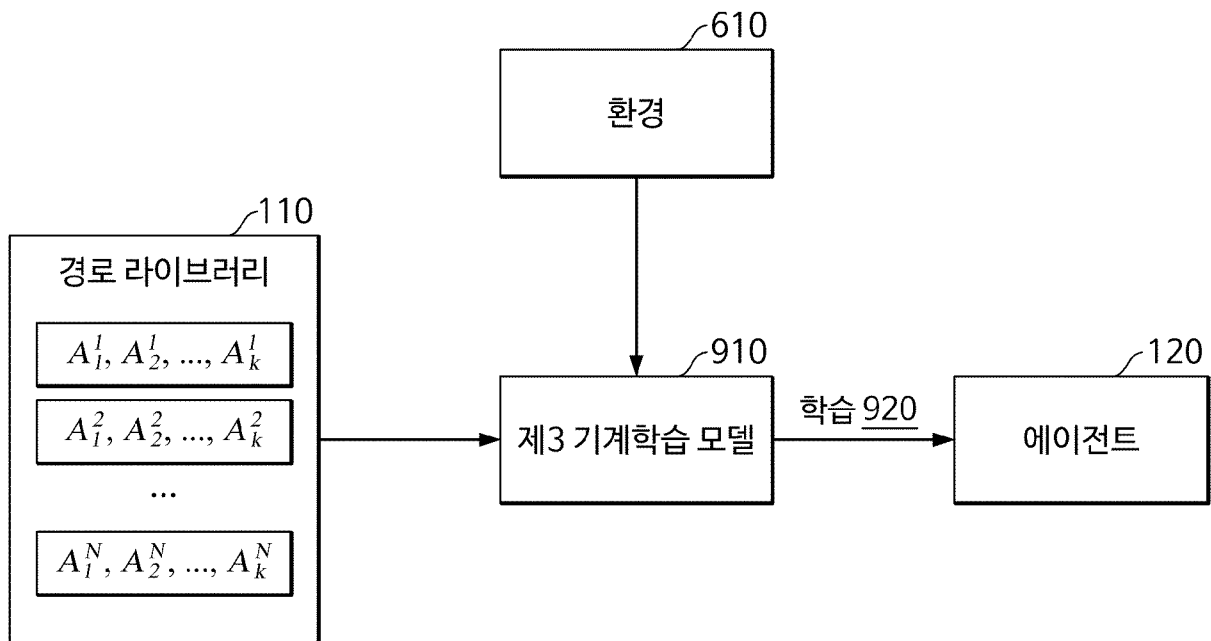
[도7]



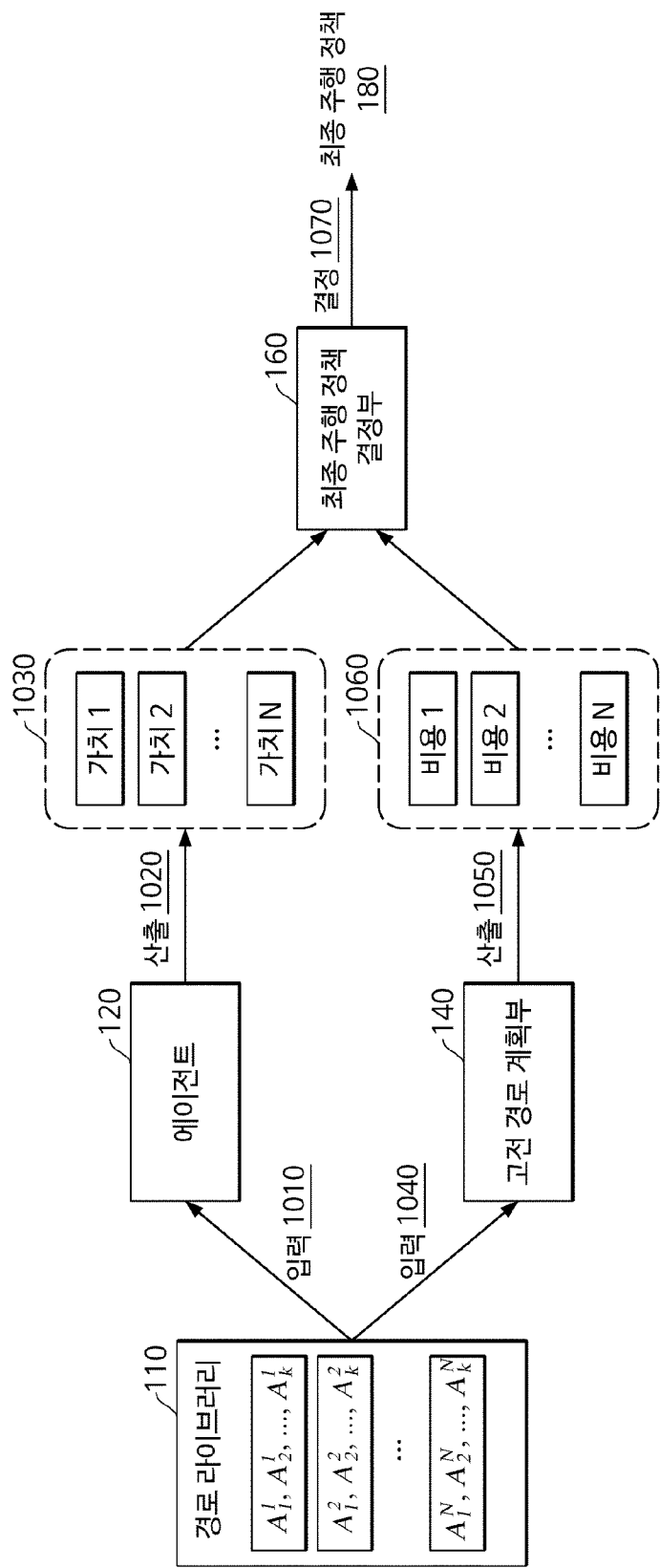
[도8]



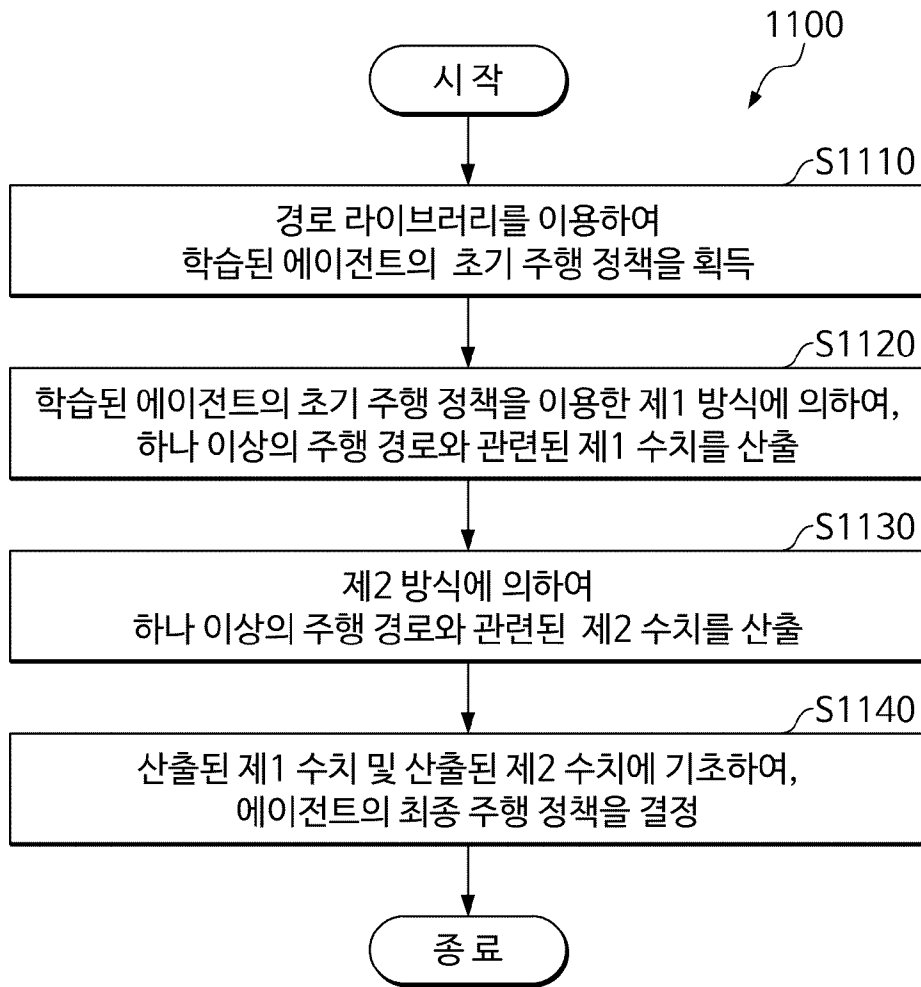
[도9]



[도10]



[도11]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2023/017342

A. CLASSIFICATION OF SUBJECT MATTER G05D 1/20(2024.01)i; G06N 20/00(2019.01)i; G01C 21/34(2006.01)i According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) G05D 1/20(2024.01); B25J 9/16(2006.01); G05D 1/02(2006.01); G06N 3/08(2006.01) Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Korean utility models and applications for utility models: IPC as above Japanese utility models and applications for utility models: IPC as above Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) eKOMPASS (KIPO internal) & keywords: 경로 라이브러리(trajecory library), 정책(policy), 보상(reward), 비용(cost), 학습(learning)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	KR 10-2021-0046500 A (ELECTRONICS AND TELECOMMUNICATIONS RESEARCH INSTITUTE) 28 April 2021 (2021-04-28) See paragraphs [0013]-[0072].	1-20
A	KR 10-2020-0010982 A (INDUSTRY-ACADEMIC COOPERATION FOUNDATION, KUNSAN NATIONAL UNIVERSITY) 31 January 2020 (2020-01-31) See paragraphs [0041]-[0128].	1-20
A	KR 10-2022-0141646 A (ELECTRONICS AND TELECOMMUNICATIONS RESEARCH INSTITUTE) 20 October 2022 (2022-10-20) See claims 1-13.	1-20
A	KR 10-2018-0053142 A (HYUNDAI MOTOR COMPANY) 21 May 2018 (2018-05-21) See claims 1-20.	1-20
A	KR 10-2019-0109624 A (SAMSUNG ELECTRONICS CO., LTD.) 26 September 2019 (2019-09-26) See claims 1-18.	1-20
☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.		
* Special categories of cited documents: “A” document defining the general state of the art which is not considered to be of particular relevance “D” document cited by the applicant in the international application “E” earlier application or patent but published on or after the international filing date “L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) “O” document referring to an oral disclosure, use, exhibition or other means “P” document published prior to the international filing date but later than the priority date claimed “T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention “X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone “Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art “&” document member of the same patent family		
Date of the actual completion of the international search 02 February 2024		Date of mailing of the international search report 02 February 2024
Name and mailing address of the ISA/KR Korean Intellectual Property Office Government Complex-Daejeon Building 4, 189 Cheongsaro, Seo-gu, Daejeon 35208 Facsimile No. +82-42-481-8578		Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/KR2023/017342

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
KR	10-2021-0046500	A	28 April 2021	US	11772758	B2	03 October 2023
				US	2021-0114698	A1	22 April 2021
KR	10-2020-0010982	A	31 January 2020	KR	10-2124553	B1	18 June 2020
KR	10-2022-0141646	A	20 October 2022	None			
KR	10-2018-0053142	A	21 May 2018	CN	108072382	A	25 May 2018
				DE	102017205518	A1	17 May 2018
				KR	10-2518532	B1	07 April 2023
				US	10073458	B2	11 September 2018
				US	2018-0136662	A1	17 May 2018
KR	10-2019-0109624	A	26 September 2019	CN	111566578	A	21 August 2020
				EP	3695281	A1	19 August 2020
				KR	10-2481487	B1	27 December 2022
				US	11047698	B2	29 June 2021
				US	2019-0265060	A1	29 August 2019
				WO	2019-168336	A1	06 September 2019

A. 발명이 속하는 기술분류(국제특허분류(IPC))

G05D 1/20(2024.01)i; G06N 20/00(2019.01)i; G01C 21/34(2006.01)i

B. 조사된 분야

조사된 최소문헌(국제특허분류를 기재)

G05D 1/20(2024.01); B25J 9/16(2006.01); G05D 1/02(2006.01); G06N 3/08(2006.01)

조사된 기술분야에 속하는 최소문헌 이외의 문헌

한국등록실용신안공보 및 한국공개실용신안공보: 조사된 최소문헌란에 기재된 IPC

일본등록실용신안공보 및 일본공개실용신안공보: 조사된 최소문헌란에 기재된 IPC

국제조사에 이용된 전산 데이터베이스(데이터베이스의 명칭 및 검색어(해당하는 경우))

eKOMPASS(특허청 내부 검색시스템) & 키워드: 경로 라이브러리(trajecory library), 정책(policy), 보상(reward), 비용(cost), 학습(learning)

C. 관련 문헌

카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
A	KR 10-2021-0046500 A (한국전자통신연구원) 2021.04.28 단락 [0013]-[0072]	1-20
A	KR 10-2020-0010982 A (군산대학교산학협력단) 2020.01.31 단락 [0041]-[0128]	1-20
A	KR 10-2022-0141646 A (한국전자통신연구원) 2022.10.20 청구항 1-13	1-20
A	KR 10-2018-0053142 A (현대자동차주식회사) 2018.05.21 청구항 1-20	1-20
A	KR 10-2019-0109624 A (삼성전자주식회사) 2019.09.26 청구항 1-18	1-20



추가 문헌이 C(계속)에 기재되어 있습니다.



대응특허에 관한 별지를 참조하십시오.

* 인용된 문헌의 특별 카테고리:

“A” 특별히 관련이 없는 것으로 보이는 일반적인 기술수준을 정의한 문헌

“D” 본 국제출원에서 출원인이 인용한 문헌

“E” 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일 이후에 공개된 선출원 또는 특허 문헌

“L” 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의 공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여 인용된 문헌

“O” 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌

“P” 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌

“T” 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과 상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기 위해 인용된 문헌

“X” 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의 신규성 또는 진보성이 없는 것으로 본다.

“Y” 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른 문헌과 조합하는 경우로 그 조합이 당업자에게 자명한 경우 청구된 발명은 진보성이 없는 것으로 본다.

“&” 동일한 대응특허문헌에 속하는 문헌

국제조사의 실제 완료일

2024년02월02일(02.02.2024)

국제조사보고서 발송일

2024년02월02일(02.02.2024)

ISA/KR의 명칭 및 우편주소

대한민국 특허청
(35208) 대전광역시 서구 청사로 189, 4동 (둔산동,
정부대전청사)

팩스 번호 +82-42-481-8578

심사관

양정록

전화번호 +82-42-481-5709

국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
KR 10-2021-0046500 A	2021/04/28	US 11772758 B2	2023/10/03
		US 2021-0114698 A1	2021/04/22
KR 10-2020-0010982 A	2020/01/31	KR 10-2124553 B1	2020/06/18
KR 10-2022-0141646 A	2022/10/20	없음	
KR 10-2018-0053142 A	2018/05/21	CN 108072382 A	2018/05/25
		DE 102017205518 A1	2018/05/17
		KR 10-2518532 B1	2023/04/07
		US 10073458 B2	2018/09/11
		US 2018-0136662 A1	2018/05/17
KR 10-2019-0109624 A	2019/09/26	CN 111566578 A	2020/08/21
		EP 3695281 A1	2020/08/19
		KR 10-2481487 B1	2022/12/27
		US 11047698 B2	2021/06/29
		US 2019-0265060 A1	2019/08/29
		WO 2019-168336 A1	2019/09/06