



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2019년12월12일

(11) 등록번호 10-2055141

(24) 등록일자 2019년12월06일

(51) 국제특허분류(Int. Cl.)

H04L 12/751 (2013.01) H04L 12/727 (2013.01)

H04L 12/801 (2013.01) H04Q 9/00 (2006.01)

(52) CPC특허분류

H04L 45/08 (2013.01)

H04L 45/121 (2013.01)

(21) 출원번호 10-2018-0173606

(22) 출원일자 2018년12월31일

심사청구일자 2018년12월31일

(56) 선행기술조사문헌

KR1020140063392 A*

KR1020190086133 A

*는 심사관에 의하여 인용된 문헌

(73) 특허권자

한국기술교육대학교 산학협력단

충청남도 천안시 동남구 병천면 충절로 1600 (한국기술교육대학교내)

(72) 발명자

한연희

대전광역시 유성구 가정로 306-6, 101동 1201호
(도룡 SK VIEW)

임현교

대전광역시 중구 태평로 65, 35동 96호 (삼부아파트)

(뒷면에 계속)

(74) 대리인

김견수

전체 청구항 수 : 총 8 항

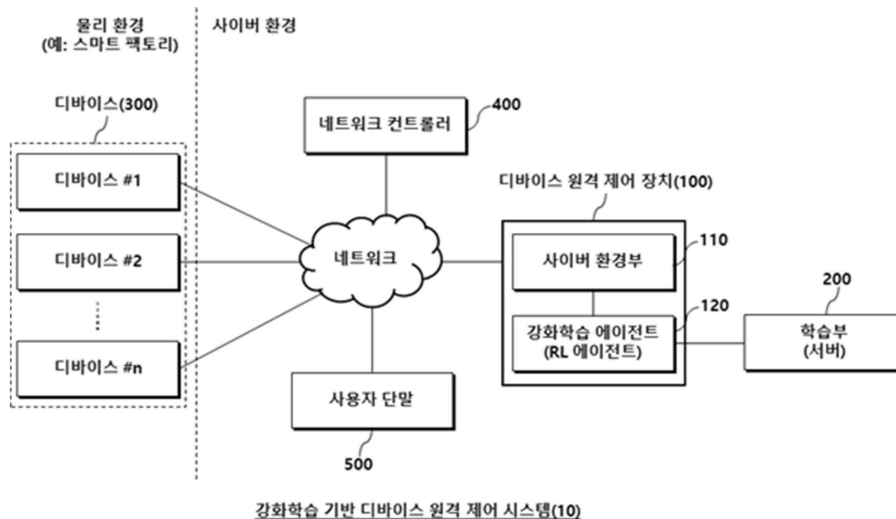
심사관 : 윤태섭

(54) 발명의 명칭 강화학습 기반 디바이스 원격 제어 시스템 및 그 방법

(57) 요약

본 발명은 강화학습 기반 디바이스 원격 제어 시스템 및 그 방법에 관한 것으로, 물리 환경에 위치하는 복수의 디바이스를 제어하기 위한 학습모델을 구축하고, 상기 구축한 학습모델을 통해 사이버 환경에서 상기 디바이스를 제어하기 위한 제어명령을 생성하여 상기 디바이스에 제공한 후, 상기 각 디바이스로부터 상기 제어명령에 따른 상태정보를 수신하여, 상기 수신한 상태정보에 따라 상기 학습모델에 대한 강화학습을 수행함으로써, 상기 학습모델을 고도화시킴과 동시에 상기 상태정보와 제어명령에 대해서는 네트워크 지연시간을 최소화하여 원격에서 상기 디바이스를 실시간으로 정밀하게 제어할 수 있도록 하는 강화학습 기반 디바이스 원격 제어 시스템 및 그 방법에 관한 것이다.

대표도 - 도1



(52) CPC특허분류

H04L 47/29 (2013.01)

H04Q 9/00 (2013.01)

(72) 발명자

김주봉

충청북도 청주시 상당구 쇠내로126번길 40, 삼일아
파트 305-406

권도형

대전광역시 서구 계룡로320번길 119 (복음맨션)

이 발명을 지원한 국가연구개발사업

과제고유번호 2018R1A6A1A03025526

부처명 교육부

연구관리전문기관 National Research Foundation of Korea (한국연구재단)

연구사업명 대학중점연구소지원사업

연구과제명 상호작용 가상현실 기반 몰입형 교육 훈련 플랫폼

기 여 율 1/1

주관기관 한국기술교육대학교 산학협력단

연구기간 2018.06.01 ~ 2019.02.28

명세서

청구범위

청구항 1

물리 환경에 위치하는 적어도 하나 이상의 디바이스;

상기 디바이스를 제어하기 위한 제어정보를 강화학습 에이전트로부터 제공받아 상기 디바이스로 송신하고, 상기 디바이스로부터 해당 제어정보에 따라 제어된 상기 디바이스의 상태정보를 수신하고 상기 수신한 디바이스 상태 정보에 대한 리워드 정보를 생성하여 상기 디바이스의 상태정보와 상기 리워드 정보를 강화학습 에이전트로 제공하는 사이버 환경부;

상기 사이버 환경부로부터 상기 수신한 디바이스의 상태정보를 디바이스 제어용 학습모델에 적용하여 상기 제어 정보를 추출하고, 상기 추출한 제어정보를 상기 사이버 환경부로 제공하며, 상기 추출한 제어정보를 토대로 제어되기 전의 디바이스 상태정보, 상기 제어정보, 상기 제어정보를 토대로 제어된 후의 디바이스 상태정보 및 상기 제어된 후의 디바이스 상태정보에 대한 리워드 정보를 포함하는 강화학습 데이터를 생성하여 학습부로 제공하는 강화학습 에이전트; 및

디바이스 제어용 학습네트워크를 통해 사전에 정의된 강화학습 데이터를 학습하여, 상기 디바이스를 특정 상태에서 다음의 상태로 제어하기 위한 제어정보를 추출하기 위한 상기 디바이스 제어용 학습모델의 학습 파라미터를 생성하는 학습부;를 포함하며,

상기 강화학습 에이전트는, 상기 적어도 하나 이상의 디바이스별로 설정된 고유의 기능에 따라 특정 작업을 수행하는 일련의 과정인 에피소드에 따라 상기 디바이스를 제어할 때, 상기 강화학습 데이터를 누적하여 저장하고, 상기 에피소드가 종료되면, 상기 누적하여 저장한 강화학습 데이터를 미리 설정한 수만큼 랜덤하게 선택하여 배치한 다음 상기 학습부로 제공하고, 상기 학습부로부터 학습 파라미터를 제공받아 상기 디바이스 제어용 학습모델에 적용하여 강화학습을 수행하도록 하며,

상기 디바이스 제어용 학습모델은, 상기 강화학습 데이터 간의 상관관계로 인해 강화학습의 속도가 느려지는 것을 방지하지 위해 상기 상관관계를 무시하고 상기 특정 상태에서 미리 설정한 상기 디바이스의 최종 상태까지의 상기 리워드 정보의 총합이 최대가 되는 상기 제어정보를 추출하도록 학습되는 것을 특징으로 하는 강화학습 기반 디바이스 원격 제어 시스템.

청구항 2

청구항 1에 있어서,

상기 디바이스와 상기 사이버 환경부간의 상기 상태정보 및 제어정보는,

스위치, 라우터 또는 이들의 조합을 포함하는 복수의 네트워크 노드를 통해서 송수신되며,

상기 복수의 네트워크 노드는,

네트워크 컨트롤러를 통해, 상기 상태정보 및 제어정보에 대한 최대 대역폭을 할당하고, 상기 네트워크 컨트롤러에서 산출한 최단경로에 따라 상기 상태정보 및 제어정보를 최우선적으로 송수신하도록 설정됨으로써, 상기 상태정보 및 제어정보를 실시간으로 송수신하도록 제어하는 것을 특징으로 하는 강화학습 기반 디바이스 원격 제어 시스템.

청구항 3

삭제

청구항 4

청구항 1에 있어서,

상기 리워드 정보는,

상기 디바이스의 상태정보가 미리 설정한 임계범위 내에서 제어된 결과인 경우 플러스 보상값을 적용하고, 미리 설정한 임계범위를 초과하여 제어된 결과인 경우 마이너스 보상값을 적용함으로써, 생성되는 것을 특징으로 하는 강화학습 기반 디바이스 원격 제어 시스템.

청구항 5

삭제

청구항 6

청구항 1에 있어서,

상기 학습부는,

상기 강화학습 에이전트로부터 제공받은 강화학습 데이터를 이용하여 디바이스 제어용 학습네트워크에 대한 강화학습을 수행하여, 강화학습 결과에 대한 학습 파라미터를 생성하고, 상기 생성한 학습 파라미터를 상기 강화학습 에이전트로 제공하여, 상기 디바이스 제어용 학습모델에 상기 학습 파라미터를 적용할 수 있도록 함으로써, 해당 디바이스 제어용 학습모델을 업데이트하도록 하는 것을 더 포함하는 것을 특징으로 하는 강화학습 기반 디바이스 원격 제어 시스템.

청구항 7

사이버 환경부에서, 물리 환경에 위치하는 적어도 하나 이상의 디바이스를 제어하기 위한 제어정보를 강화학습 에이전트로부터 제공받아 상기 디바이스로 송신하고, 상기 디바이스로부터 해당 제어정보에 따라 제어된 상기 디바이스의 상태정보를 수신하는 단계;

상기 사이버 환경부에서, 상기 수신한 디바이스 상태정보에 대한 리워드 정보를 생성하여 상기 디바이스의 상태정보와 상기 리워드 정보를 강화학습 에이전트로 제공하는 단계;

상기 강화학습 에이전트에서, 상기 사이버 환경부로부터 상기 수신한 디바이스의 상태정보를 디바이스 제어용 학습모델에 적용하여 상기 제어정보를 추출하고, 상기 추출한 제어정보를 상기 사이버 환경부로 제공하는 단계;

상기 강화학습 에이전트에서, 상기 추출한 제어정보를 토대로 제어되기 전의 디바이스 상태정보, 상기 제어정보, 상기 제어정보를 토대로 제어된 후의 디바이스 상태정보 및 상기 제어된 후의 디바이스 상태정보에 대한 리워드 정보를 포함하는 강화학습 데이터를 생성하여 학습부로 제공하는 단계; 및

디바이스 제어용 학습네트워크를 통해 사전에 정의된 강화학습 데이터를 학습하여, 상기 디바이스를 특정 상태에서 다음의 상태로 제어하기 위한 제어정보를 추출하기 위한 상기 디바이스 제어용 학습모델의 학습 파라미터를 생성하는 학습 단계;를 포함하며,

상기 강화학습 에이전트는, 상기 적어도 하나 이상의 디바이스별로 설정된 고유의 기능에 따라 특정 작업을 수행하는 일련의 과정인 에피소드에 따라 상기 디바이스를 제어할 때, 상기 강화학습 데이터를 누적하여 저장하고, 상기 에피소드가 종료되면, 상기 누적하여 저장한 강화학습 데이터를 미리 설정한 수만큼 랜덤하게 선택하여 배치한 다음 상기 학습부로 제공하고, 상기 학습부로부터 학습 파라미터를 제공받아 상기 디바이스 제어용 학습모델에 적용하여 강화학습을 수행하도록 하며,

상기 디바이스 제어용 학습모델은, 상기 강화학습 데이터 간의 상관관계로 인해 강화학습의 속도가 느려지는 것을 방지하기 위해 상기 상관관계를 무시하고 상기 특정 상태에서 미리 설정한 상기 디바이스의 최종 상태까지의 상기 리워드 정보의 총합이 최대가 되는 상기 제어정보를 추출하도록 학습되는 것을 특징으로 하는 강화학습 기반 디바이스 원격 제어 방법.

청구항 8

청구항 7에 있어서,

상기 디바이스와 상기 사이버 환경부간의 상기 상태정보 및 제어정보는,

스위치, 라우터 또는 이들의 조합을 포함하는 복수의 네트워크 노드를 통해서 송수신되며,

상기 복수의 네트워크 노드는,

네트워크 컨트롤러를 통해, 상기 상태정보 및 제어정보에 대한 최대 대역폭을 할당하고, 상기 네트워크 컨트롤러에서 산출한 최단경로에 따라 상기 상태정보 및 제어정보를 최우선적으로 송수신하도록 설정됨으로써, 상기 상태정보 및 제어정보를 실시간으로 송수신하도록 제어하는 것을 특징으로 하는 강화학습 기반 디바이스 원격 제어 방법.

청구항 9

청구항 7에 있어서,

상기 리워드 정보는,

상기 디바이스의 상태정보가 미리 설정한 임계범위 내에서 제어된 결과인 경우 플러스 보상값을 적용하고, 미리 설정한 임계범위를 초과하여 제어된 결과인 경우 마이너스 보상값을 적용함으로써, 생성되는 것을 특징으로 하는 강화학습 기반 디바이스 원격 제어 방법.

청구항 10

청구항 7에 있어서,

상기 학습 단계는,

상기 강화학습 에이전트로부터 제공받은 강화학습 데이터를 이용하여 디바이스 제어용 학습네트워크에 대한 강화학습을 수행하여, 강화학습 결과에 대한 학습 파라미터를 생성하고, 상기 생성한 학습 파라미터를 상기 강화학습 에이전트로 제공하여, 상기 디바이스 제어용 학습모델에 상기 학습 파라미터를 적용할 수 있도록 함으로써, 해당 디바이스 제어용 학습모델을 업데이트하도록 하는 것을 더 포함하는 것을 특징으로 하는 강화학습 기반 디바이스 원격 제어 방법.

발명의 설명

기술 분야

[0001] 본 발명은 강화학습 기반 디바이스 원격 제어 시스템 및 그 방법에 관한 것으로, 더욱 상세하게는 물리 환경에 위치하는 적어도 하나 이상의 디바이스를 제어하기 위한 학습모델을 구축하고, 상기 구축한 학습모델을 통해 사이버 환경에서 상기 디바이스를 제어하기 위한 제어명령을 생성하여 상기 디바이스에 제공한 후, 상기 각 디바이스로부터 상기 제어명령에 따른 상태정보를 수신하여, 상기 수신한 상태정보에 따라 상기 학습모델에 대한 강화학습을 수행함으로써, 상기 학습모델을 고도화시킴과 동시에 상기 상태정보와 제어명령에 대해서는 네트워크 지연시간을 최소화하여 원격에서 상기 디바이스를 실시간으로 정밀하게 제어할 수 있도록 하는 강화학습 기반 디바이스 원격 제어 시스템 및 그 방법에 관한 것이다.

배경 기술

[0002] 최근 산업기술 및 정보통신 기술의 급격한 발전으로 인해, ICT 분야에서는 머신러닝을 기반으로 하는 4차 산업혁명이 일어났으며, 머신러닝을 적용한 다양한 IoT 기기들이 개발되고 보급되고 있다.

[0003] 이로 인해 복수의 센서와 액추에이터를 포함하는 디바이스와 상기 각 디바이스를 제어하기 위한 제어시스템이 유무선 네트워크로 연결되어 사용자에게 다양한 기능을 제공하고 있다.

[0004] 그러나 종래의 제어시스템은, 상기 각 디바이스가 위치하는 로컬에 상기 각 디바이스마다 각각 구축되며, 상기 디바이스를 관리하는 관리자 혹은 사용자가 상기 제어시스템을 직접적으로 조작하여 상기 디바이스를 제어하도록 구성된다.

[0005] 이는, 상기 디바이스의 기능에 따라 고도의 정밀성을 요구하는 경우, 해당 디바이스의 전문적인 지식을 가지는 별도의 전문가가 필요하며, 상기 각 디바이스마다 제어시스템을 구축해야하기 때문에, 해당 제어시스템에 대한 구축비용이 많이 소요되는 문제점이 있다.

[0006] 이러한 문제점을 해결하기 위해 최근에는, 상기 디바이스와 제어시스템을 네트워크를 통해 연결하여 원격에서 상기 디바이스를 제어하기 위한 원격 제어시스템이 개발되어 상용화되고 있다.

[0007] 종래의 원격 제어시스템은, 상기 디바이스를 제어하기 위한 제어명령을 생성하고, 상기 네트워크를 통해 전송하

며, 상기 디바이스에서 상기 제어명령을 수신 받아 상기 수신한 제어명령에 따라 해당 디바이스를 동작하도록 구현된다.

[0008] 즉, 종래의 원격 제어시스템은, 다수의 센서와 액추에이터 등을 포함하여 구성되는 디바이스와 네트워크로 연결되어 실제 디바이스에 대한 물리적 정보를 습득, 가공, 계산 및 분석하여, 이에 대한 결과를 토대로 제어명령을 생성하고, 상기 생성한 제어명령을 상기 네트워크를 통해 전송함으로써, 상기 디바이스를 원격에서 제어할 수 있도록 한다.

[0009] 그러나 종래의 원격 제어시스템은, 네트워크를 통해 상기 제어명령을 상기 각 디바이스로 전송하는 과정에 있어서, 상기 제어명령에 대한 네트워크 대역폭을 랜덤하게 배분하고, 상기 네트워크상에 위치하는 복수의 스위치 혹은 라우터에서 상기 제어명령에 대한 경로를 각각 계산하여 송수신하기 때문에 상기 제어명령에 대한 수신 지연이 필연적으로 발생할 수밖에 없다. 이는 물리 환경에 위치하는 복수의 디바이스를 실시간으로 제어하지 못하는 문제점을 내포하고 있다.

[0010] 따라서 본 발명에서는, 물리 환경에 위치하는 복수의 디바이스를 제어하기 위해 생성한 각각의 학습모델을 통해 상기 디바이스를 제어하기 위한 제어명령을 추출하여 상기 디바이스로 제공하고, 상기 디바이스로부터 해당 제어명령에 따른 동작을 수행한 결과인 상기 디바이스의 상태정보를 수신하여, 상기 수신한 상태정보를 이용하여 상기 학습모델에 대한 강화학습을 수행함으로써, 상기 학습모델을 지속적으로 고도함과 동시에, 네트워크 컨트롤러를 통해 네트워크 트래픽을 중앙에서 관리하여 상기 상태정보 및 제어명령에 대한 우선권을 조정하고, 상기 상태정보와 제어명령에 대한 최적의 경로를 산출하여, 상기 산출한 최적의 경로를 상기 네트워크에 위치하는 라우터 혹은 스위치에 제공함으로써, 상기 상태정보 및 제어명령을 실시간으로 송수신할 수 있도록 하여 상기 디바이스를 원격에서 실시간으로 정밀하게 제어할 수 있도록 하는 방안을 제안하고자 한다.

[0011] 다음으로 본 발명의 기술분야에 존재하는 선행기술에 대하여 간단하게 설명하고, 이어서 본 발명이 상기 선행기술에 비해서 차별적으로 이루고자 하는 기술적 사항에 대해서 기술하고자 한다.

[0012] 먼저 한국공개특허 제2018-0102870호(2018.09.18.)는 전자 장치 및 그 전자 장치의 제어 방법에 관한 것으로, 특정 공간에 위치하는 사용자의 생체 정보, 사용자의 행동 정보, 해당 공간의 이용 내역을 나타내는 로그 정보, 해당 공간을 함께 이용하는 동반자 정보 등을 포함하는 사용자의 특성 정보를 학습하여, 상기 특정 공간에 대한 이용 목적을 예측하여, 상기 특정 공간에 설치된 복수의 전자 장치를 제어할 수 있도록 하는 전자 장치 및 그 전자 장치의 제어 방법에 관한 것이다.

[0013] 상기 선행기술은, 인공지능 방법을 통해 특정 공간에 위치하는 복수의 사용자에게 대한 특정 정보의 조합을 분석하여, 상기 복수의 사용자가 특정 공간을 어떠한 용도로 이용할지를 예측함으로써, 해당 공간에 설치되는 복수의 전자 장치(예: 프로젝터, 마이크, 조명, 온도와 습도를 제어하는 공조 장치 등)를 제어할 수 있도록 하는 것이다.

[0014] 반면에 본 발명은, 특정 디바이스를 제어하기 위한 학습모델을 통해 상기 특정 디바이스의 기능에 대한 에피소드에 따라 상기 특정 디바이스를 제어하고, 상기 제어한 결과를 토대로 상기 학습모델에 대한 강화학습을 수행함으로써, 다양한 환경에 설치되는 상기 디바이스를 해당 환경에 적응하여 정밀하게 제어할 수 있도록 하며, 네트워크상에서 상기 디바이스 제어에 필요한 관련 데이터를 송수신할 때, 해당 데이터에 대한 네트워크 트래픽의 우선권을 조정하여 실시간으로 송수신함으로써, 상기 디바이스에 대한 실시간 제어가 가능하도록 하는 것이다.

[0015] 따라서 상기 선행기술의 기술적 특징과 본 발명의 기술적 특징은 현저한 차이점이 있다.

[0016] 또한 한국등록특허 제10-0768564호(2007.10.12.)는 원격지 디바이스 제어 시스템 및 그 제어 방법에 관한 것으로, 사용자 클라이언트로부터 원격지 디바이스의 정보를 수집하거나, 제어하기 위한 명령데이터를 수신하고, CDMA, TDMA, 이동통신망을 이용하여 원격지 디바이스로 상기 명령데이터를 전송함으로써, 상기 원격지 디바이스로부터 디바이스의 정보를 수집하거나, 제어할 수 있도록 하는 원격지 디바이스 제어 시스템 및 그 제어 방법에 관한 것이다.

[0017] 상기 선행기술은 이동통신사의 ISP 망을 통해 상기 명령데이터를 송수신하는 것으로, 상기 ISP 망에서 명령데이터의 송수신에 대한 지연이 발생하는 경우, 상기 원격지 디바이스를 실시간으로 제어할 수 없는 문제점이 있다.

[0018] 반면에 본 발명은, 네트워크를 중앙에서 제어 및 관리하기 위한 네트워크 컨트롤러를 통해 모든 네트워크 트래픽의 플로우 테이블(flow-table)을 관리하며, 디바이스의 상태정보와 해당 디바이스의 고유 기능에 따른 에피소드에 따라 상기 디바이스를 제어하기 위한 제어명령을 상기 네트워크를 통해 송수신할 때, 상기 상태정보와 상기

제어명령에 대한 네트워크 트래픽 플로우의 우선권을 조정하고, 상기 상태정보 및 제어명령에 대한 최적 경로를 산출하여, 상기 산출한 최적 경로대로 상기 상태정보 및 제어명령을 라우팅 또는 스위칭할 수 있도록 함으로써, 상기 디바이스를 원격에서 실시간으로 제어할 수 있도록 하는 것이다.

[0019] 따라서, 상기 선행기술은 본 발명의 이러한 기술적 특징을 기재하거나 시사하고 있지 않다.

발명의 내용

해결하려는 과제

[0020] 본 발명은 상기와 같은 문제점을 해결하기 위해 창안된 것으로서, 물리적 환경에 위치하는 복수의 디바이스를 사이버 환경에서 원격으로 자동 제어할 수 있는 디바이스 제어용 학습모델을 각각 생성하여, 상기 각 디바이스의 고유기능에 대한 에피소드에 따라 정확하고 효율적으로 제어할 수 있도록 하는 강화학습 기반 디바이스 원격 제어 시스템 및 그 방법을 제공하는 것을 그 목적으로 한다.

[0021] 또한 본 발명은, 상기 디바이스 제어용 학습모델을 통해 상기 디바이스를 제어하기 위한 제어명령을 추출하고, 상기 추출한 제어명령을 네트워크를 통해 상기 디바이스로 제공함으로써, 상기 디바이스를 원격에서 제어할 수 있도록 하는 강화학습 기반 디바이스 원격 제어 시스템 및 그 방법을 제공하는 것을 또 다른 목적으로 한다.

[0022] 또한 본 발명은, 상기 디바이스로부터 상기 제어명령을 수행한 결과인 해당 디바이스의 상태정보를 수신하고, 상기 수신한 상태정보를 토대로 상기 생성한 디바이스 제어용 학습모델에 대한 강화학습을 수행하여 상기 디바이스 제어용 학습모델을 지속적으로 고도화함으로써, 상기 물리 환경에 적응하여 상기 디바이스를 정밀하게 제어할 수 있도록 하는 강화학습 기반 디바이스 원격 제어 시스템 및 그 방법을 제공하는 것을 또 다른 목적으로 한다.

[0023] 이때, 상기 강화학습 기반 디바이스 원격 제어 시스템은, 상기 에피소드가 종료될 때 까지, 상기 수신한 상태정보를 상기 제어명령에 따른 제어결과에 대한 리워드를 계산하고, 상기 제어명령을 제공하기 전의 디바이스의 상태(현재 상태)와 상기 제어명령, 상기 계산한 리워드 및 상기 제어명령을 수행한 디바이스의 상태(다음 상태)를 매핑함으로써, 강화학습을 위한 강화학습 데이터를 누적하여 저장하고, 상기 저장한 강화학습 데이터를 상기 학습모델에 적용함으로써, 상기 강화학습을 수행하도록 구현된다.

[0024] 또한 본 발명은, 네트워크 컨트롤러를 통해 상기 상태정보 및 상기 제어명령에 대한 네트워크상에서의 트래픽이 최대 우선권을 가지도록 조정함으로써, 네트워크 지연 없이 상기 디바이스를 실시간으로 제어할 수 있도록 하는 상기 디바이스를 우선권을 최대 우선권을 조정함으로써, 강화학습 기반 디바이스 원격 제어 시스템 및 그 방법을 제공하는 것을 또 다른 목적으로 한다.

과제의 해결 수단

[0025] 본 발명의 일 실시예에 따른 강화학습 기반 디바이스 원격 제어 시스템은, 물리 환경에 위치하는 적어도 하나 이상의 디바이스, 상기 디바이스를 제어하기 위한 제어정보를 상기 디바이스로 송신하고, 상기 디바이스로부터 해당 제어정보에 따라 제어된 상기 디바이스의 상태정보를 상기 디바이스로부터 수신하는 사이버 환경부 및 상기 수신한 디바이스의 상태정보를 디바이스 제어용 학습모델에 적용하여 상기 제어정보를 추출하고, 상기 추출한 제어정보를 상기 사이버 환경부로 제공하는 강화학습 에이전트를 포함하는 것을 특징으로 한다.

[0026] 또한 상기 상태정보 및 제어정보는, 스위치, 라우터 또는 이들의 조합을 포함하는 복수의 네트워크 노드를 통해서 송수신되며, 상기 복수의 네트워크 노드는, 네트워크 컨트롤러를 통해, 상기 상태정보 및 제어정보에 대한 최대 대역폭을 할당하고, 상기 네트워크 컨트롤러에서 산출한 상기 상태정보 및 제어정보에 대한 최단경로에 따라 최우선적으로 송수신하도록 설정됨으로써, 상기 상태정보 및 제어정보를 실시간으로 송수신하도록 제어되는 것을 특징으로 한다.

[0027] 또한 상기 디바이스 원격 제어 시스템은, 디바이스 제어용 학습네트워크를 통해 사전에 정의된 학습데이터를 학습하여, 상기 디바이스를 특정 상태에서 다음의 상태로 제어하기 위한 제어정보를 추출하기 위한 디바이스 제어용 학습모델을 생성하는 학습부를 더 포함하고, 상기 학습데이터는, 상기 디바이스의 현재 상태정보, 제어정보, 리워드 정보 및 상기 제어정보를 토대로 제어된 다음 상태정보를 포함하며, 상기 학습모델은, 상기 특정 상태에서 미리 설정한 상기 디바이스의 최종 상태까지의 리워드 정보의 총합이 최대가 되는 상기 제어정보를 추출하도록 학습되는 것을 특징으로 한다.

- [0028] 또한 상기 사이버 환경부는, 상기 수신한 디바이스 상태정보에 대한 리워드 정보를 생성하는 것을 더 포함하며, 상기 리워드 정보는, 상기 디바이스의 상태정보가 미리 설정한 임계범위 내에서 제어된 결과인 경우 플러스 보상값을 적용하고, 미리 설정한 임계범위를 초과하여 제어된 결과인 경우 마이너스 보상값을 적용함으로써, 생성되는 것을 특징으로 한다.
- [0029] 또한 상기 강화학습 에이전트는, 상기 사이버 환경부로부터 디바이스 상태정보 및 해당 디바이스 상태정보에 대한 리워드 정보를 수신하여, 상기 수신한 디바이스 상태정보 및 해당 디바이스 상태정보에 대한 리워드 정보를 토대로 강화학습 데이터를 생성하여, 학습부로 제공하는 것을 더 포함하며, 상기 강화학습 데이터는, 상기 제어정보를 토대로 제어되기 전의 디바이스 상태정보, 해당 제어정보, 상기 제어정보를 토대로 제어된 후의 디바이스 상태정보 및 상기 제어된 후의 디바이스 상태정보에 대한 리워드 정보를 포함하는 것을 특징으로 한다.
- [0030] 또한 상기 학습부는, 상기 강화학습 에이전트로부터 제공받은 강화학습 데이터를 이용하여 디바이스 제어용 학습네트워크에 대한 강화학습을 수행하여, 강화학습 결과에 대한 학습 파라미터를 생성하고, 상기 생성한 학습 파라미터를 상기 강화학습 에이전트로 제공하여, 상기 디바이스 제어용 학습모델에 상기 학습 파라미터를 적용할 수 있도록 함으로써, 해당 디바이스 제어용 학습모델을 업데이트하도록 하는 것을 더 포함하는 것을 특징으로 한다.
- [0031] 아울러 본 발명의 일 실시예에 따른 강화학습 기반 디바이스 원격 제어 방법은, 물리 환경에 위치하는 적어도 하나 이상의 상기 디바이스를 제어하기 위한 제어정보를 상기 디바이스로 송신하고, 상기 디바이스로부터 해당 제어정보에 따라 제어된 상기 디바이스의 상태정보를 상기 디바이스로부터 수신하는 단계, 상기 수신한 디바이스의 상태정보를 디바이스 제어용 학습모델에 적용하여 상기 제어정보를 추출하고, 상기 추출한 제어정보를 상기 디바이스로 제공하는 단계를 포함하는 것을 특징으로 한다.
- [0032] 또한 상기 디바이스 원격 제어 방법은, 사전에 정의된 학습데이터를 학습하여, 상기 디바이스를 특정 상태에서 다음의 상태로 제어하기 위한 제어정보를 추출하기 위한 디바이스 제어용 학습모델을 생성하는 학습 단계를 더 포함하고, 상기 학습데이터는, 상기 디바이스의 현재 상태정보, 제어정보, 리워드 정보 및 상기 제어정보를 토대로 제어된 다음 상태정보를 포함하며, 상기 학습모델은, 상기 특정 상태에서 미리 설정한 상기 디바이스의 최종 상태까지의 리워드 정보의 총합이 최대가 되는 상기 제어정보를 추출하도록 학습되는 것을 특징으로 한다.
- [0033] 또한 상기 강화학습 기반 디바이스 원격 제어 방법은, 상기 수신한 디바이스 상태정보에 대한 리워드 정보를 생성하는 단계, 상기 수신한 디바이스 상태정보 및 해당 디바이스 상태정보에 대해 생성한 상기 리워드 정보를 토대로 강화학습 데이터를 생성하는 단계 및 상기 생성한 강화학습 데이터를 이용하여 디바이스 제어용 학습네트워크에 대한 강화학습을 수행함으로써, 강화학습 결과에 대한 학습 파라미터를 생성하고, 상기 생성한 학습 파라미터를 상기 디바이스 제어용 학습모델에 적용하여 해당 디바이스 제어용 학습모델을 업데이트하는 단계를 더 포함하는 것을 특징으로 한다.

발명의 효과

- [0034] 이상에서와 같이 본 발명의 강화학습 기반 디바이스 원격 제어 시스템 및 그 방법은, 다양한 환경에 노출되는 복수의 디바이스를 원격에서 제어하기 위한 디바이스 제어용 학습모델을 상기 디바이스를 제어한 제어결과에 따른 디바이스 상태정보를 토대로 강화학습을 수행하여 고도화함으로써, 상기 다양한 환경에 적응하여 상기 디바이스를 정밀하고 효과적으로 제어할 수 있도록 하는 효과가 있다.
- [0035] 또한 중앙의 네트워크 컨트롤러에서 네트워크상의 모든 트래픽에 대한 플로우 테이블을 관리할 수 있도록 하며, 상기 디바이스 제어용 학습모델을 통해 추출되는 제어명령과 해당 제어명령에 따른 상기 디바이스의 상태정보를 송수신할 때, 상기 네트워크 컨트롤러를 통해 상기 제어명령 및 상기 상태정보의 트래픽에 대한 우선권을 조정하고, 상기 제어명령 및 상기 상태정보에 대한 네트워크상의 최적 경로를 산출하여, 상기 산출한 최적 경로에 따라 상기 제어명령 및 상태정보를 송수신할 수 있도록 함으로서, 상기 디바이스를 실시간으로 제어할 수 있도록 하는 효과가 있다.

도면의 간단한 설명

- [0036] 도 1은 본 발명의 일 실시예에 따른 강화학습 기반 디바이스 원격 제어 시스템 및 그 방법을 개략적으로 설명하기 위해 나타낸 도면이다.
- 도 2는 본 발명의 일 실시예에 따른 네트워크 컨트롤러의 기능을 설명하기 위해 나타낸 도면이다.

도 3은 본 발명의 일 실시예에 따른 디바이스 원격 제어 장치를 통해 물리 환경에 위치하는 복수의 디바이스를 제어하는 방법을 설명하기 위해 나타낸 도면이다.

도 4는 본 발명의 일 실시예에 따른 디바이스 제어용 학습네트워크의 구조를 나타낸 도면이다.

도 5는 본 발명의 일 실시예에 따른 강화학습 기반 디바이스 원격 제어 시스템의 사이버 환경부에 대한 구성을 나타낸 블록도이다.

도 6은 본 발명의 일 실시예에 따른 강화학습 기반 디바이스 원격 제어 시스템의 강화학습 에이전트에 대한 구성을 나타낸 블록도이다.

도 7은 본 발명의 일 실시예에 따른 강화학습 기반 디바이스 원격 제어 시스템의 학습부에 대한 구성을 나타낸 블록도이다.

도 8은 본 발명의 일 실시예에 따른 강화학습 기반 디바이스 원격 제어 방법에 대한 절차를 나타낸 흐름도이다.

발명을 실시하기 위한 구체적인 내용

- [0037] 이하, 첨부한 도면을 참조하여 본 발명의 강화학습 기반 디바이스 원격 제어 시스템 및 그 방법에 대한 바람직한 실시예를 상세히 설명한다. 각 도면에 제시된 동일한 참조부호는 동일한 부재를 나타낸다. 또한 본 발명의 실시예들에 대해서 특정한 구조적 내지 기능적 설명들은 단지 본 발명에 따른 실시예를 설명하기 위한 목적으로 예시된 것으로, 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가지고 있다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 의미를 가지는 것으로 해석되어야 하며, 본 명세서에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는 것이 바람직하다.
- [0038] 도 1은 본 발명의 일 실시예에 따른 강화학습 기반 디바이스 원격 제어 시스템 및 그 방법을 개략적으로 설명하기 위해 나타낸 도면이다.
- [0039] 도 1에 도시한 바와 같이, 본 발명의 일 실시예에 따른 강화학습 기반 디바이스 원격 제어 시스템(10)은, 물리적 환경에 위치하는 복수의 디바이스(300), 상기 복수의 디바이스(300)를 사이버 환경에서 원격으로 제어하기 위한 디바이스 원격 제어 장치(100), 상기 디바이스(300)에 대한 원격 제어명령을 생성하여 상기 디바이스를 제어할 수 있도록 하는 디바이스 제어용 학습모델을 생성하고, 상기 생성한 학습모델에 대한 강화학습을 수행하는 학습부(200), 네트워크를 통해 상기 디바이스 원격 제어 장치(100)와 상기 디바이스(300)간 데이터를 송수신하기 위해 상기 네트워크의 트래픽을 관리 및 조정하는 네트워크 컨트롤러(400) 및 사용자 단말(500)을 포함하여 구성된다.
- [0040] 또한 상기 복수의 디바이스(300)는, 스마트 팩토리, 팩내 등과 같이 다양한 물리적 환경에 위치하고, 모터(motor), 액추에이터(actuator), 특정 정보를 수집하기 위한 다양한 센서 등을 구비하여 고유의 기능을 수행하기 위한 장치를 의미한다.
- [0041] 예를 들어, 상기 디바이스(300)는, 특정 물건을 미리 설정한 위치로 옮기기 위한 로봇 팔(robot arm), 또는 특정 공간을 자동으로 청소하기 위한 청소 로봇 등과 같은 다양한 장치일 수 있다. 즉, 상기 디바이스(300)는 미리 정해진 에피소드에 따라 그 고유의 기능을 수행하는 다양한 장치를 의미하는 것이다.
- [0042] 또한 디바이스 원격 제어 장치(100)는, 네트워크를 통해 상기 디바이스(300)와 상호 인터랙션을 수행하여 상기 각 디바이스(300)로부터 실제 물리적인 정보를 수집하고, 상기 수집한 실제 물리적인 정보를 토대로 상기 디바이스(300)를 사이버 환경에서 원격으로 제어할 수 있도록 하는 기능을 수행한다.
- [0043] 한편 상기 네트워크는, 5G(fifth generation), 이더넷(ethernet) 등과 같은 주지의 유무선 통신네트워크를 의미한다.
- [0044] 또한 상기 디바이스 원격 제어 장치(100)는, 상기 디바이스(300)를 제어하기 위한 제어정보를 상기 디바이스로(300)로 전송하고, 상기 제어정보에 따른 상기 디바이스의 실제 물리적인 정보인 상태정보를 수신하는 사이버 환경부(110) 및 상기 학습부(200)를 통해 생성된 학습모델로부터 상기 디바이스(300)를 제어하기 위한 제어명령을 추출하여 상기 사이버 환경부(110)를 통해 상기 디바이스(300)로 전송할 수 있도록 하는 강화학습 에이전트(120)를 포함하여 구성된다.

- [0045] 상기 학습모델은 상기 디바이스 원격 제어 장치(100)와 연결되고, 서버로 구축되는 학습부(200)에서 생성되어 상기 강화학습 에이전트(120)로 제공된다.
- [0046] 즉, 상기 학습부(200)는 사전에 생성된 학습데이터를 학습하여, 상기 디바이스(300)를 원격에서 제어할 수 있도록 제어정보를 추출하기 위한 학습모델을 생성하여, 상기 강화학습 에이전트(120)로 제공한다.
- [0047] 이후, 상기 학습부(200)는, 상기 강화학습 에이전트(120)로부터 제공되는 강화학습 데이터를 토대로 상기 생성한 학습모델에 대한 강화학습을 지속적으로 수행하여, 해당 학습모델을 고도화하는 기능을 수행한다.
- [0048] 한편 상기 학습데이터는, 사전에 생성되어 데이터베이스(미도시)에 저장된 것으로, 상기 강화학습 기반 디바이스 원격 제어 시스템(10)의 관리자 혹은 사용자가 별도의 제어기(미도시) 또는 상기 디바이스 원격 제어 장치(100)를 통해 상기 디바이스(300)를 수동으로 제어하여, 그 결과를 누적함으로써 생성될 수 있다.
- [0049] 다만 상기 학습데이터는, 상기 디바이스(300)가 제조되어 출하될 때, 해당 디바이스(300)의 제조자에 의한 테스트를 통해 생성될 수 도 있다. 즉, 본 발명에서 상기 학습데이터를 생성하는 방법에 대해서는 그 제한을 두지 아니한다.
- [0050] 또한 상기 학습데이터는, 상기 디바이스(300)에 대한 현재 상태정보(state), 제어정보(action), 상기 제어정보를 토대로 상기 디바이스(300)가 제어된 상태인 다음의 상태정보(next state), 상기 제어정보를 토대로 상기 현재 상태정보에서 다음의 상태정보로 전이될 때, 미리 설정한 임계 범위 내에서 상기 디바이스(300)가 적절하게 잘 동작되었는지에 대한 여부를 나타내는 리워드(reward) 정보를 포함한다.
- [0051] 즉, 상기 학습데이터는, 상기 제어정보에 따라 제어되기 전의 상기 디바이스(300)에 대한 현재 상태정보, 상기 제어정보, 상기 제어정보에 따라 제어된 후에 해당하는 상기 디바이스(300)에 대한 다음 상태정보, 상기 다음 상태정보에 대한 리워드를 포함하는 것이다.
- [0052] 또한 상기 사이버 환경부(110)는, 미리 설정된 에피소드에 따라 상기 물리 환경에 위치하는 특정 디바이스(300)를 제어하고자 하는 경우, 상기 사이버 환경부(110)를 통해 상기 특정 디바이스(300)를 초기화하며, 상기 특정 디바이스(300)로부터 해당 디바이스(300)에 대한 상태정보를 전송받는다.
- [0053] 이때, 상기 사이버 환경부(110)는, 현재 전송받은 상태정보가 제어정보에 따라 상기 디바이스(300)가 제어된 결과인 경우에는, 해당 상태정보에 대한 리워드를 계산하여 상기 강화학습 에이전트(120)로 제공하고, 상기 전송받은 상태정보가 상기 초기화에 따른 최초 상태정보인 경우에는, 상기 리워드를 계산하지 않고 해당 상태정보를 상기 강화학습 에이전트(120)로 제공하게 된다.
- [0054] 또한 상기 강화학습 에이전트(120)는, 상기 사이버 환경부(110)로부터 제공받은 디바이스(300)의 상태정보를 토대로 강화학습 데이터를 생성하여 강화학습 에이전트(120)의 메모리(미도시)에 저장한다.
- [0055] 상기 강화학습 데이터는, 상기 디바이스(300)의 현재 상태정보, 제어정보, 리워드 및 다음 상태정보를 포함하는 것으로, 이때, 상기 현재 상태정보는 상기 제어정보를 토대로 제어되기 전의 디바이스(300)의 상태정보를 의미하며, 상기 다음 상태정보는 상기 제어정보를 토대로 제어되어 특정 상태로 전이된 상태의 디바이스(300)에 대한 상태정보를 의미한다. 또한 상기 리워드는, 상기 사이버 환경부(110)에서 계산하여 제공되는 것으로, 상기 제어정보를 토대로 제어된 상기 디바이스(300)에 대한 다음 상태정보에 대한 보상값을 의미한다.
- [0056] 또한 상기 강화학습 데이터는, 상기 에피소드가 종료될 때까지 누적되어 상기 데이터베이스에 저장되며, 상기 학습부(200)는 상기 누적되어 저장된 강화학습 데이터를 토대로 상기 생성한 학습모델에 대한 강화학습을 수행한다.
- [0057] 즉, 상기 강화학습은 상기 에피소드가 종료될 때마다, 수행되는 것이며, 상기 강화학습에 대해서는 도 3을 참조하여 상세히 설명하도록 한다.
- [0058] 한편 상기 현재 상태정보 및 상기 다음 상태정보는, 상기 디바이스(300)를 동작시키기 위해 구비되는 모터의 각도, 각속도, 또는 상기 디바이스(300)의 위치정보 등을 포함할 수 있다.
- [0059] 즉, 상기 현재 상태정보 및 상기 다음 상태정보는, 상기 디바이스(300)의 종류에 따라 다양하게 설정될 수 있다.
- [0060] 또한 상기 제어정보 또한 상기 디바이스(300)의 종류에 따라 다양하게 설정될 수 있으며, 미리 설정된 에피소드에 따라 상기 디바이스(300)를 구동시키기 위해 필요한 전력량, 연료량 등을 포함할 수 있다.

- [0061] 또한 상기 리워드는, 상기 제어정보에 따라 미리 설정된 임계범위 내에서 상기 디바이스(300)가 적절하게 동작하였는지에 대한 여부에 따라 미리 설정값 값으로 부여된다.
- [0062] 예를 들어, 상기 디바이스(300)가 임의의 물건을 특정 위치로 옮기는 기능을 수행하는 로봇 팔인 경우, 상기 현재 상태정보 및 다음의 상태정보는, 해당 로봇 팔에 구비되는 모터의 각속도, 각도, 상기 로봇 팔의 각속도, 각도 또는 상기 물건의 위치정보를 포함할 수 있고, 상기 제어정보는 상기 모터를 동작시키기 위한 전력량을 포함할 수 있으며, 상기 리워드는 미리 설정한 각도 범위에서 상기 모터 및 로봇 팔이 구동된 경우에는 상기 에피소드대로 적절하게 구동된 것으로 판단하여 미리 설정한 보상값(예: +1)을 다음 상태정보에 대해 적용하고, 미리 설정한 각도 범위를 초과하여 상기 모터 및 로봇 팔이 구동된 경우에는 미리 설정한 보상값(예: -100)을 적용함으로써, 상기 리워드를 적용할 수 있다.
- [0063] 즉, 상기 초기 학습모델을 생성하기 위한 학습데이터나, 상기 학습모델을 강화하기 위한 강화학습 데이터에 포함되는 현재 상태정보, 제어정보, 리워드, 다음 상태정보는, 강화학습 기반 디바이스 원격 제어 시스템(10)을 통해 원격에서 제어하기 위한 디바이스(300)의 종류에 따라 각각 다르게 설정될 수 있다.
- [0064] 상기에서 설명한 것과 같이, 본 발명의 강화학습 기반 디바이스 원격 제어 시스템(10)은, 각 디바이스(300)별 에피소드대로 상기 디바이스(300)를 제어하는 과정을 반복적으로 수행하고, 상기 디바이스(300)를 제어한 결과에 대한 강화학습 데이터를 각 에피소드에 따라 생성하여, 상기 생성한 강화학습 데이터를 이용하여 상기 학습모델을 지속적으로 업데이트함으로써, 해당 학습모델에 대한 강화학습을 수행하여 상기 학습모델을 점진적으로 고도화한다. 이를 통해 상기 강화학습 기반 디바이스 원격 제어 시스템(10)은, 다양한 환경에 위치하는 상기 디바이스(300)를 해당 환경에 적응하여, 해당 디바이스(300)를 정밀하게 제어할 수 있도록 한다.
- [0065] 또한 네트워크 컨트롤러(400)는, 상기 디바이스 원격 제어 장치(100)를 통해 상기 디바이스(300)를 실시간으로 제어할 수 있도록, 상기 네트워크 트래픽에 대한 우선권을 조정하는 기능을 수행한다.
- [0066] 즉, 상기 네트워크 컨트롤러(400)는, 상기 네트워크를 통해 상기 디바이스(300)에서 상기 디바이스 원격 제어 장치(100)로 전송되는 상태정보와 상기 디바이스 원격 제어 장치(100)에서 상기 디바이스(300)로 전송되는 제어정보의 네트워크 트래픽에 대한 우선권을 최대 우선권으로 설정함으로써, 상기 상태정보 및 제어정보가 실시간 송수신이 가능하도록 하여, 상기 디바이스(300)를 원격에서 실시간으로 제어할 수 있도록 한다.
- [0067] 한편 상기 네트워크 컨트롤러(400)에 대해서는, 도 2를 참조하여 상세히 설명하도록 한다.
- [0068] 도 2는 본 발명의 일 실시예에 따른 네트워크 컨트롤러의 기능을 설명하기 위해 나타낸 도면이다.
- [0069] 도 2에 도시한 바와 같이, 본 발명의 일 실시예에 따른 네트워크 컨트롤러(400)는, 오픈 플로우(open flow) 프로토콜을 통해 상기 네트워크를 구성하는 복수의 네트워크 노드(600)와 통신을 수행하여, 상기 각 노드에 대한 동작을 설정하는 기능을 수행한다.
- [0070] 이때, 상기 노드(600)는, 네트워크를 통해 전송되는 데이터를 목적지로 전송하는 기능을 수행하는 것으로, 스위치(예: OVS(open virtual switch)), 라우터 등을 의미한다.
- [0071] 본 발명에서의 네트워크는 SDN(software define network) 기술이 적용되어 구성되는 것으로, 상기 각 노드(600)의 컨트롤 플레인(control plane)과 데이터 플레인(data plane)을 분리함으로써, 네트워크 세부 정보에 상관없이 중앙의 상기 네트워크 컨트롤러(400)를 통해 상기 네트워크 노드(600)를 중앙에서 제어 및 관리할 수 있도록 한다.
- [0072] 즉, 상기 네트워크 컨트롤러(400)는 상기 복수의 네트워크 노드(600)로부터 노드정보를 실시간으로 수신하여, 상기 각 노드(600)를 관리 및 제어하는 기능을 수행한다.
- [0073] 한편 상기 노드정보는, 각 네트워크 노드(600)를 통해 송수신되는 모든 데이터에 대한 네트워크 트래픽의 플로우 테이블(flow table)과 각 네트워크 노드(600)에서의 데이터 전송경로에 대한 혼잡여부에 대한 정보를 포함한다.
- [0074] 따라서, 상기 네트워크 컨트롤러(400)는, 상기 각 노드(600)로부터 실시간으로 수신되는 노드정보를 토대로 상기 각 노드(600)를 통해 송수신되는 모든 데이터에 대한 네트워크 트래픽과 상기 각 데이터의 송수신 경로상의 혼잡여부를 실시간으로 모니터링할 수 있다.
- [0075] 또한 물리적 환경에 위치하는 복수의 디바이스(300)와 사이버 환경에 위치하여 상기 각 디바이스(300)를 원격에서 제어하기 위한 디바이스 원격 제어 장치(100)는 상기 네트워크를 구성하는 복수의 네트워크 노드(600)를 통

해 상기 제어를 위한 데이터(즉, 상태정보 및 제어정보)를 송수신하게 된다.

- [0076] 한편 상기 디바이스 원격 제어 장치(100)를 통해 상기 복수의 디바이스(300)를 각각 실시간으로 제어하기 위해서는, 상기 복수의 디바이스(300)로부터 전송되는 디바이스 상태정보와 상기 디바이스 원격 제어 장치(100)로부터 전송되는 제어정보는 실시간으로 송수신되는 것이 매우 중요하다.
- [0077] 따라서 상기 네트워크 컨트롤러(400)는 상기 복수의 노드(600)로부터 수신되는 네트워크 트래픽에 대한 플로우 테이블을 토대로 네트워크 트래픽의 플로우를 모니터링하고 있으면서, 상기 상태정보와 상기 제어명령에 대한 네트워크 트래픽에 대한 대역폭을 최대 대역폭으로 할당할 수 있도록 상기 각 노드(600)를 설정한다.
- [0078] 또한 상기 네트워크 컨트롤러(400)는, 상기 상태정보와 상기 제어명령에 대한 네트워크 트래픽에 대한 우선권을 조정하여, 상기 각 네트워크 노드(600)에서 상기 상태정보와 상기 제어명령을 다른 데이터보다 최우선적으로 송수신할 수 있도록 설정한다.
- [0079] 즉, 상기 네트워크 컨트롤러(400)는, 상기 플로우 테이블을 참조하여 각 네트워크 노드(600)에 대한 네트워크 트래픽을 확인하고, 상기 각 네트워크 트래픽에 대한 우선권을 조정함으로써, 상기 상태정보 및 제어정보 이외의 트래픽에 대해서는 낮은 우선권으로 설정하고, 상태정보 및 제어정보에 대한 네트워크 트래픽에 대해서는 높은 우선권으로 설정함으로써, 상기 상태정보 및 제어정보를 우선적으로 송수신할 수 있도록 하는 것이다.
- [0080] 또한 상기 네트워크 컨트롤러(400)는, 상기 수신되는 노드정보의 데이터 전송경로에 대한 혼잡여부에 대한 정보를 참조하여, 상기 상태정보 및 제어명령에 대한 최단 경로를 산출하여, 상기 산출한 최단 경로를 상기 각 네트워크 노드(600)로 전송함으로써, 상기 각각의 네트워크 노드(600)에서 상기 최단 경로에 따라 상기 상태정보 및 제어정보를 전송할 수 있도록 한다.
- [0081] 즉, 본 발명의 네트워크 컨트롤러(400)는, 상기 상태정보 및 제어정보에 대한 최대 대역폭할당, 우선권 조정 및 최단 경로 산출을 통해 상기 디바이스(300)와 상기 디바이스 원격 제어 장치(100) 간의 상기 상태정보 및 제어정보를 실시간으로 송수신할 수 있도록 함으로써, 상기 디바이스 원격 제어 장치(100)를 통해 상기 디바이스(300)를 원격에서 실시간으로 제어할 수 있도록 한다.
- [0082] 한편 상기 네트워크 컨트롤러(400)는, 상기 각 노드정보와 네트워크 트래픽에 대한 우선권정보를 REST API(representational state transfer API)를 이용하여 상기 사용자 단말(500)로 제공함으로써, 사용자로 하여금 상기 노드정보 및 우선권정보를 포함하는 네트워크 상태를 실시간으로 모니터링할 수 있도록 한다.
- [0083] 또한 상기 사용자는 상기 사용자 단말(500)을 통해 상기 상태정보 및 제어정보에 대한 네트워크 트래픽의 우선권을 조정할 수 있으며, 상기 우선권을 조정한 우선권 조정정보를 상기 REST API를 통해 상기 네트워크 컨트롤러(400)로 제공함으로써, 상기 조정한 우선권을 상기 상태정보 및 제어정보에 적용할 수 있도록 한다.
- [0084] 도 3은 본 발명의 일 실시예에 따른 디바이스 원격 제어 장치를 통해 물리 환경에 위치하는 복수의 디바이스를 제어하는 방법을 설명하기 위해 나타낸 도면이다.
- [0085] 도 3에 도시한 바와 같이, 본 발명의 일 실시예에 따른 디바이스 원격 제어 장치(100)는, 사이버 환경부(110) 및 강화학습 에이전트(120)를 포함하며, 상기 복수의 디바이스(300)와 학습부(200)와 연동하여, 다양한 환경에 노출되는 상기 각각의 디바이스(300)를 원격에서 제어하는 기능을 수행한다.
- [0086] 상기 디바이스(300)를 원격에서 제어하기 위해, 본 발명의 디바이스 원격 제어 시스템(10)은, 상기 학습부(200)를 통해 상기 디바이스(300)별로 미리 설정된 에피소드에 따른 제어정보를 추출하기 위한 학습모델(즉, 디바이스 제어용 학습모델)을 생성한다.
- [0087] 상기 생성되는 학습모델은, 디바이스 제어용 학습네트워크를 통해 사전에 미리 생성된 초기 학습데이터를 학습함으로써, 생성되는 것으로, 상기 각각의 디바이스(300)별로 생성된다.
- [0088] 한편 상기 초기 학습데이터는, 상기 디바이스 원격 제어 시스템(10)의 사용자 또는 관리자에 의해 생성되는 것으로, 도 1을 참조하여 설명하였으므로, 더 이상의 상세한 설명은 생략하도록 한다.
- [0089] 또한 상기 학습부(200)를 통해 생성된 학습모델은, 상기 강화학습 에이전트(120)로 제공되며, 상기 강화학습 에이전트(120)는 상기 학습부(200)에서 제공되는 학습모델을 자체적으로 구비되는 메모리(미도시)에 저장하여, 상기 저장한 학습모델로부터 상기 디바이스(300)를 제어하기 위한 제어정보를 추출하게 된다.
- [0090] 상기 디바이스 원격 제어 장치(100)는, 미리 설정된 에피소드대로 상기 디바이스(300)를 제어하고자 할 때 우선

적으로, 상기 사이버 환경부(110)를 통해 상기 디바이스(300)를 초기화한다.

- [0091] 이 때, 상기 사이버 환경부(110)는, 상기 디바이스(300)를 초기화하기 위한 제어정보를 상기 디바이스(300)로 송신함으로써, 상기 디바이스(300)를 초기화할 수 있도록 하며, 상기 디바이스(300)는, 상기 초기화를 수행한 후, 디바이스 상태정보를 상기 사이버 환경부(110)로 제공한다.
- [0092] 또한 상기 사이버 환경부(110)는, 상기 디바이스(300)로부터 수신 받은 디바이스 상태정보가 초기화를 위한 제어정보에 의한 디바이스 상태정보인 경우, 상기 강화학습 에이전트(120)로 전송한다.
- [0093] 이 때, 상기 강화학습 에이전트(120)는, 상기 수신 받은 디바이스 상태정보를 상기 메모리에 저장한 학습모델에 입력하여, 상기 디바이스(300)를 제어하기 위한 제어정보를 생성하여, 상기 사이버 환경부(110)를 통해 상기 디바이스(300)로 전송할 수 있도록 한다.
- [0094] 이 후, 상기 디바이스(300)는, 상기 수신한 제어정보를 토대로 해당 디바이스(300)를 동작시키고, 이에 대한 디바이스 상태정보를 생성하여, 상기 사이버 환경부(110)로 전송한다.
- [0095] 또한 상기 사이버 환경부(110)는, 상기 수신한 디바이스 상태정보가 상기 초기화를 위한 제어정보가 아닌 상기 강화학습 에이전트(120)에서 생성한 제어정보를 토대로 생성된 경우에는, 해당 디바이스 상태정보에 대한 리워드를 계산함으로써, 해당 디바이스 상태정보에 대한 리워드 정보를 생성하여, 상기 생성한 리워드 정보와 상기 수신한 해당 디바이스 상태정보를 상기 강화학습 에이전트(120)로 제공한다.
- [0096] 상기 상태정보 및 제어정보에 대한 송수신은, 임의의 특정 통신방법을 이용하여 수행되며, 상기 상태정보 및 제어정보는 상기 통신방법에 적합한 데이터 포맷으로 변환되어 송수신된다.
- [0097] 예를 들면, 상기 통신방법은, 통신기기 사이에서 통신할 수 있는 경량의 발행(publish), 구독(subscribe) 메시징 전송 프로토콜을 의미하는 것으로, M2M(machine to machine), IoT(internet of things) 등에 최적화되어 낮은 전력 및 낮은 대역폭 환경에서 고속으로 통신을 수행하는 것을 포함한다.
- [0098] 또한 상기 강화학습 에이전트(120)는, 상기 수신한 디바이스 상태정보 및 리워드 정보를 토대로 상기 학습모델을 강화하기 위한 강화학습 데이터를 생성하여 상기 메모리에 저장한다.
- [0099] 상기 강화학습 데이터는, 상기 에피소드가 종료될 때 까지 누적되어 저장된다.
- [0100] 한편 상기 에피소드는, 상기 각 디바이스(300)로별 고유의 기능에 따라 특정 작업을 수행하는 일련의 과정을 의미하는 것으로, 예를 들어, 상기 디바이스(300)가 로봇 팔인 경우, 해당 로봇 팔을 통해 특정 객체를 특정 위치로 이동시키기 위해 상기 디바이스(300)를 제어할 때, 상기 특정 객체를 특정 위치로 이동시키는 일련의 과정이 하나의 에피소드가 된다.
- [0101] 한편 상기 강화학습 데이터는, 상기 디바이스(300)의 현재 상태정보, 제어정보, 상기 디바이스(300)의 다음 상태정보에 대한 리워드 정보 및 상기 디바이스(300)의 다음 상태정보를 포함하여 구성된다.
- [0102] 또한 상기 디바이스(300)의 현재 상태정보는, 상기 제어정보를 토대로 해당 디바이스(300)가 동작되기 전의 상태정보를 의미하며, 상기 디바이스(300)의 다음 상태정보는, 상기 제어정보를 토대로 해당 디바이스(300)가 동작된 후의 상태정보를 의미한다.
- [0103] 또한 상기 리워드 정보는, 상기 제어정보를 토대로 제어된 후의 다음 상태정보가 미리 설정한 임계범위를 기준으로 주어지는 보상값을 의미하는 것으로, 상기 다음 상태정보가 미리 설정한 임계범위 내에서 동작한 결과인 경우 플러스 보상값 (예: +1), 상기 다음 상태정보가 미리 설정한 임계범위를 초과하여 동작한 결과인 경우 마이너스 보상값(예: -100)으로 부여될 수 있다.
- [0104] 또한 상기 강화학습 에이전트(120)는, 상기 에피소드에 따라 상기 디바이스(300)를 제어하는 것이 종료된 경우, 상기 누적하여 저장한 강화학습 데이터를 상기 학습부(200)로 제공하여, 상기 디바이스 제어용 학습네트워크에 대한 강화학습을 수행할 수 있도록 한다.
- [0105] 또한 상기 학습부(200)는, 상기 강화학습 에이전트(120)로부터 수신한 강화학습 데이터를 이용하여, 상기 디바이스 제어용 학습네트워크에 대한 강화학습을 수행하고, 이에 대한 결과인 강화학습 파라미터를 상기 강화학습 에이전트(120)로 제공한다.
- [0106] 이때, 상기 강화학습은, 상기 리워드 정보를 토대로 상기 디바이스(300)의 현재 상태정보에서 다음 상태정보로 동작하기 위한 제어정보를 추출할 때, 상기 에피소드가 종료될 때까지의 전체 리워드 정보가 제일 높은 다음 상

태정보로 제어하기 위한 제어정보가 추출되도록 수행된다.

- [0107] 즉, 상기 강화학습은, 디바이스(300)의 현재 상태에서 특정 제어명령에 따라 다음 상태로 전이될 때, 다음 상태에 전이시키기 위한 제어명령에 대한 리워드 정보가 제일 높은 제어명령을 추출하는 것이 아니라, 현재 상태에서 시나리오가 종료될 때 까지 전체 리워드 정보가 높은 다음 상태정보에 대한 제어명령을 추출하게 되는 것이다.
- [0108] 다시 말해, 상기 학습부(200)는, 상기 강화학습 에이전트(120)에서 제공한 강화학습 데이터를 토대로, 특정 상태정보가 입력되는 경우, 전체 리워드 정보가 높은 제어정보를 출력하도록 상기 디바이스 제어용 학습네트워크에 대한 강화학습을 수행하게 된다.
- [0109] 한편 상기 디바이스 제어용 학습네트워크는 CNN(convolucional neural network)을 이용하여 구축되는 것이 바람직하지만, ANN(artificial neural network)등과 같은 다양한 기계학습 네트워크를 통해 구축될 수 있다. 즉, 본 발명의 일 실시예에 따른 디바이스 제어용 학습네트워크에 대한 구조는 그 제한을 두지 아니한다.
- [0110] 또한 상기 강화학습 결과에 대한 학습 파라미터를 수신한 강화학습 에이전트(120)는, 상기 수신한 학습 파라미터를 상기 메모리에 저장된 학습모델에 적용함으로써, 해당 학습모델을 업데이트하며, 이를 통해 상기 디바이스 원격 제어를 위한 학습모델을 고도화한다.
- [0111] 이러한 강화학습은, 상기 에피소드가 끝날 때 마다 반복적으로 수행되며, 이러한 과정을 통해 상기 학습모델을 지속적으로 업데이트함으로써, 다양한 환경에 노출되는 각각의 디바이스(300)를 상기 환경에 적응하여 정밀하게 제어할 수 있도록 한다.
- [0112] 따라서, 상기 강화학습 에이전트(120)는, 임의의 환경에 위치하는 디바이스(300)의 현재 상태를 인식하여, 선택 가능한 제어정보 중 전체 에피소드에서 리워드 정보의 합이 최대인 제어정보를 선택함으로써, 상기 디바이스(300)를 정밀하게 제어할 수 있도록 한다.
- [0113] 도 4는 본 발명의 일 실시예에 따른 디바이스 제어용 학습네트워크의 구조를 나타낸 도면이다.
- [0114] 도 4에 도시한 바와 같이, 본 발명의 일 실시예에 따른 디바이스 제어용 학습네트워크는, 상기 강화학습 에이전트(120)에서 생성한 강화학습 데이터를 입력받는 입력 레이어, 미리 설정한 크기의 커널(kernel)을 이용하여 상기 커널과 상기 강화학습 데이터의 특정 부분을 컨볼루션(convolution)하는 제1 컨볼루션 레이어(convolution layer)와 제2 컨볼루션 레이어와 두 개의 완전연관레이어(fully connected layer) 및 출력레이어로 구성된다.
- [0115] 상기 입력 레이어는, 상기 강화학습 에이전트(120)에서 각 에피소드마다 생성한 복수의 강화학습 데이터를 입력받으며, 상기 강화학습 에이전트(120)는, 상기 생성한 강화학습 데이터를 상기 디바이스 제어용 학습네트워크에 적합한 이미지로 변환하여 상기 학습부(200)로 제공한다.
- [0116] 이때, 상기 강화학습 데이터는, 강화학습 에이전트(120)에서 생성되어 저장된 모든 강화학습 데이터를 상기 학습부(200)로 제공되는 것이 아니라, 미리 설정한 수만큼 랜덤하게 선택하여 배치하고, 상기 배치한 강화학습 데이터를 이미지화하여 상기 학습부(200)로 제공 된다.
- [0117] 한편 상기 강화학습 데이터를 미리 설정한 수만큼 랜덤하게 선택하여 배치하는 이유는, 상기 강화학습 데이터에 포함된 현재 상태정보 및 다음 상태정보가 제어정보에 따라 순차적으로 전이되는 과정을 포함하고 있기 때문에, 상기 각 강화학습 데이터 간의 상관관계(correlation)로 인해 상기 강화학습에 대한 학습속도가 매우 느려지는 문제점이 발생할 수 있으므로, 상기 상관관계를 무시하고, 특정 상태정보에서 상기 에피소드가 종료될 때까지의 리워드 정보의 총합이 제일 높은 제어명령을 추출할 수 있도록 신속하게 강화학습이 수행되도록 하기 위함이다.
- [0118] 또한 상기 제1 컨볼루션 레이어는, 미리 설정한 크기(예: 4x4)와 특정 가중치를 가지는 제1 커널을 미리 설정한 스트라이드의 크기(예: 1)에 따라 이동해 가면서 상기 이미지로 변환한 강화학습 데이터의 특정 부분과 상기 제1 커널의 가중치를 각각 컨볼루션하여, 상기 이미지의 특정 부분에 대한 제1 특징 맵을 생성하여 출력한다. 여기서, 상기 스트라이드의 크기는 상기 커널의 이동단위를 의미한다.
- [0119] 이때, 상기 제1 컨볼루션 레이어를 컨볼루션을 통해 출력되는 결과는, 7x1의 크기를 가지는 총 32개(즉, 7x1x32)의 특징 맵이며, 상기 제2 컨볼루션은, 상기 제1 컨볼루션 레이어를 통해 출력되는 제1 특징 맵을 미리 설정한 크기(예: 4x1)와 특정 가중치를 가지는 제2 커널을 미리 설정한 스트라이드의 크기(예: 1)에 따라 이동해 가면서 상기 제1 컨볼루션 레이어에서 출력한 제1 특징 맵으로부터 제2 특징 맵을 생성하여 출력한다. 제2 컨볼루션 레이어의 제2 커널의 크기가 4x1인 경우, 상기 제2 컨볼루션 레이어를 통해 출력되는 제2 특징 맵은

4x1의 크기를 가지는 총 64개의 특징 맵으로 구성될 수 있다.

- [0120] 다음으로 상기 두 개의 완전연관레이어는, 상기 출력한 특징 맵을 연결시켜 상기 디바이스(300)의 특정 상태에서 다음 상태로 전이시키기 위한 적어도 하나 이상의 제어명령을 출력 레이어를 통해 출력할 수 있도록 한다.
- [0121] 이때 상기 제어명령은, 상기 특정 상태에서 다음 상태로 상기 디바이스를 전이시킬 때, 해당 디바이스(300)에 설정된 에피소드가 종료되는 종료 상태까지 각 상태에 대한 리워드 정보의 총합이 높은 순서대로 적어도 하나 이상 출력된다.
- [0122] 이러한 과정을 통해 상기 학습부(200)는 상기 에피소드에 따라 상기 디바이스(300)를 제어할 때마다 생성되는 강화학습 데이터를 강화학습데이터를 상기 디바이스 제어용 학습네트워크에 적용함으로써, 상기 디바이스 제어용 학습네트워크를 점진적으로 업데이트할 수 있다.
- [0123] 이를 통해, 상기 학습부(200)는 상기 생성한 학습모델을 지속적으로 고도화할 수 있도록 하며, 다양한 환경에 위치하는 디바이스(300)를 해당 환경에 적응하여 원격에서 정밀하게 제어할 수 있도록 한다.
- [0124] 도 5는 본 발명의 일 실시예에 따른 강화학습 기반 디바이스 원격 제어 시스템의 사이버 환경부에 대한 구성을 나타낸 블록도이다.
- [0125] 도 5에 도시한 바와 같이, 본 발명의 일 실시예에 따른 강화학습 기반 디바이스 원격 제어 시스템(10)은 디바이스 원격 제어 장치(100), 서버 형태로 구성되는 학습부(200), 물리 환경에 위치하는 디바이스(300), 네트워크 컨트롤러(400) 등을 포함하여 구성된다.
- [0126] 상기 디바이스 원격 제어 장치(100)는, 상기 디바이스 원격 제어 장치(100)와 상기 디바이스(300) 간의 원격 제어를 위한 관련 데이터를 송수신하기 위한 사이버 환경부(110) 및 상기 디바이스(300)를 원격에서 제어하기 위한 제어정보를 생성하는 강화학습 에이전트(120)를 포함하여 구성된다.
- [0127] 상기 사이버 환경부(110)는, 상기 강화학습 에이전트(120)에서 생성한 제어정보를 상기 디바이스(300)로 전송하기 위한 제어정보 전송부(111), 상기 전송한 제어정보에 따라 상기 디바이스(300)의 상태가 전이된 상태정보를 수신하는 디바이스 상태정보 수신부(112), 상기 수신한 상태정보에 대한 리워드 정보를 생성하는 리워드 정보 생성부(113) 및 통신부(114)를 포함하여 구성된다.
- [0128] 상기 제어정보 전송부(111)는, 미리 설정된 에피소드에 따라 상기 디바이스(300)를 제어하기 위한 제어정보를 상기 강화학습 에이전트(120)로부터 제공받아, 상기 통신부(114)를 통해 상기 디바이스(300)로 전송하는 기능을 수행한다.
- [0129] 한편 상기 제어정보 전송부(111)는, 상기 디바이스(300)를 제어하고자 할 때, 상기 디바이스(300)를 초기화하기 위한 제어정보를 생성하여 상기 디바이스(300)로 전송하는 기능을 더 포함한다.
- [0130] 또한 상기 디바이스 상태정보 수신부(112)는, 상기 초기화를 위한 제어정보 또는 상기 강화학습 에이전트(120)에서 생성한 제어정보에 따라 상기 디바이스(300)가 구동되어 상기 디바이스(300)의 새로운 상태로 전이된 디바이스 상태정보를 상기 디바이스(300)로부터 전송되어 상기 통신부(114)를 통해 수신하여, 상기 강화학습 에이전트(120)로 제공하는 기능을 수행한다.
- [0131] 한편 상기 제어정보 및 상태정보는, 임의의 특정 통신방법을 이용하여 송수신되며, 상기 통신부(114)는, 상기 제어정보를 특정 통신방법에 적합한 포맷으로 변환하여 상기 디바이스(300)로 전송하거나, 상기 특정 포맷으로 수신되는 디바이스 상태정보를 디코딩하여 상기 디바이스 상태정보 수신부(112)로 제공한다.
- [0132] 또한 상기 리워드 정보 생성부(113)는, 상기 수신한 디바이스 상태정보를 토대로 미리 설정한 임계범위를 기준으로 해당 디바이스 상태정보에 대한 리워드를 계산함으로써, 상기 디바이스 상태정보에 대한 리워드 정보를 생성하여, 상기 강화학습 에이전트(120)로 제공하는 기능을 수행한다.
- [0133] 상기 리워드 정보는, 상기 수신한 디바이스 상태정보가 상기 임계범위 내에서 동작된 경우에는 플러스 보상값을 적용하며, 상기 수신한 디바이스 상태정보가 미리 설정한 임계범위를 초과하여 동작된 경우에는 마이너스 보상값을 적용함으로써, 생성된다.
- [0134] 또한 강화학습 에이전트(120)는, 상기 사이버 환경부(110)에서 제공받은 디바이스 상태정보를 메모리(미도시)에 저장된 학습모델에 적용하여, 해당 상태정보를 토대로 상기 디바이스(300)를 제어하기 위한 제어정보를 생성하며, 상기 생성한 제어정보를 상기 사이버 환경부(110)의 제어정보 전송부(111)를 통해 상기 디바이스(300)로 전

송할 수 있도록 함으로써, 상기 디바이스(300)를 제어할 수 있도록 한다.

- [0135] 또한 강화학습 에이전트(120)는, 상기 사이버 환경부(110)에서 제공받은 디바이스 상태정보 및 해당 디바이스 상태정보에 대한 리워드 정보를 토대로 상기 학습부(200)에서, 디바이스 제어용 학습네트워크에 대한 강화학습을 수행하기 위한 기반이 되는 강화학습 데이터를 생성하여 메모리에 저장한다.
- [0136] 상기 강화학습 데이터는, 상기 디바이스(300)의 현재 상태정보, 제어정보, 리워드 정보 및 다음 상태정보를 포함하여 구성된다.
- [0137] 즉, 상기 강화학습 에이전트(120)는, 상기 사이버 환경부(110)에서 제공받은 디바이스 상태정보 및 리워드 정보를 다음 상태정보 및 리워드 정보로 설정하고, 상기 다음 상태정보의 이전 상태정보를 현재 상태정보로 설정하며, 상기 이전 상태정보에 대해 생성한 제어정보를 상기 강화학습 데이터의 제어정보로 설정함으로써, 상기 강화학습 데이터를 생성하게 되는 것이다.
- [0138] 이러한 강화학습 데이터는, 상기 설정한 에피소드가 종료될 때까지 누적되어 저장된다. 한편 상기 설정한 에피소드가 종료될 때는, 상기 디바이스(300)의 제어가 최종적으로 완료되는 상태를 나타내는 것으로, 상기 제어에 따라 상기 디바이스(300)가 미리 설정된 종료 상태로 최종적으로 전이된 때를 의미한다.
- [0139] 이후, 상기 강화학습 에이전트(120)는, 상기 저장한 복수의 강화학습 데이터 중 미리 설정한 개수만큼 랜덤하게 선택하여 상기 학습부(200)로 제공함으로써, 상기 디바이스 제어용 학습네트워크에 대한 강화학습을 수행할 수 있도록 한다.
- [0140] 또한 상기 강화학습 에이전트(120)는, 상기 학습부(200)에서 강화학습을 수행한 결과인 학습 파라미터를 상기 학습부(200)로부터 제공받아, 상기 제공받은 학습 파라미터를 메모리에 저장된 학습모델에 적용함으로써, 해당 학습모델을 업데이트한다.
- [0141] 도 6은 본 발명의 일 실시예에 따른 강화학습 기반 디바이스 원격 제어 시스템의 강화학습 에이전트에 대한 구성을 나타낸 블록도이다.
- [0142] 도 6에 도시한 바와 같이, 본 발명의 일 실시예에 따른 강화학습 기반 디바이스 원격 제어 시스템(10)에 구비되는 강화학습 에이전트(120)는, 상기 사이버 환경부(110)로부터 수신되는 디바이스 상태정보를 토대로 상기 디바이스(300)를 제어하기 위한 제어정보를 생성하는 제어정보 생성부(121), 상기 사이버 환경부(110)로부터 수신되는 디바이스 상태정보 및 해당 디바이스 상태정보에 대한 리워드 정보를 토대로 디바이스 원격 제어를 위한 학습모델에 대한 강화학습을 수행하기 위한 토대가 되는 강화학습 데이터를 생성하는 강화학습 데이터 생성부(122), 학습모델 업데이트부(123) 및 상기 학습모델 및 강화학습 데이터를 저장하기 위한 메모리(124)를 포함하여 구성된다.
- [0143] 상기 제어정보 생성부(121)는, 상기 사이버 환경부(110)로부터 제공되는 디바이스 상태정보를 상기 메모리(124)에 저장한 학습모델에 입력하여, 상기 디바이스 상태정보를 토대로 해당 디바이스(300)를 제어하기 위한 제어정보를 생성하는 기능을 수행한다.
- [0144] 상기 제어정보 생성부(121)는, 해당 디바이스(300)에 대한 학습모델을 상기 메모리(124)로부터 로딩하고, 상기 로딩한 학습모델에 상기 제공받은 디바이스 상태정보를 입력하여, 해당 디바이스(300)를 제어하기 위한 제어정보를 생성하게 된다.
- [0145] 이때, 상기 학습모델은, 미리 설정한 디바이스(300)의 에피소드에 따라 최종 상태까지의 리워드 정보의 총합이 높은 순으로 적어도 하나 이상의 제어정보를 출력하며, 상기 제어정보 생성부(121)는, 상기 리워드 정보가 제일 높은 제어정보를 선택함으로써, 상기 디바이스(300)를 제어하기 위한 제어정보를 생성하게 된다.
- [0146] 또한 상기 강화학습 데이터 생성부(122)는, 상기 사이버 환경부(110)에서 제공받은 디바이스 상태정보 및 해당 디바이스 상태정보에 대한 리워드 정보를 토대로 상기 학습부(200)의 디바이스 제어용 학습네트워크에 대한 강화학습을 수행하기 위한 학습데이터인 강화학습 데이터를 생성하는 기능을 수행한다.
- [0147] 상기 강화학습 데이터는 상기 미리 설정한 에피소드가 종료될 때 까지, 상기 제어정보에 따라 수신되는 디바이스 상태정보 및 리워드 정보를 토대로 누적되어 생성되고, 상기 메모리(124)에 저장됨은 상술한 바와 같다.
- [0148] 또한 상기 강화학습 데이터 생성부(122)는, 상기 설정한 에피소드가 종료되어 상기 디바이스(300)에 대한 제어가 종료된 경우, 상기 저장된 복수의 강화학습 데이터를 랜덤 미니 배치(random mini batch)를 통해 미리 설정한 개수만큼 선택하고, 상기 선택한 강화학습 데이터를 이미지화하여 상기 학습부(200)로 제공하여, 상기 학습

부(200)를 통해 상기 디바이스 제어용 학습네트워크에 대한 강화학습을 수행할 수 있도록 한다.

- [0149] 한편 상기 강화학습 데이터를 이미지화하여 제공하는 것은, 상기 디바이스 제어용 학습네트워크가 CNN을 기반으로 구현된 경우에 수행되는 것으로, CNN 이외에 다른 기계학습 네트워크를 기반으로 상기 디바이스 제어용 학습네트워크가 구현되는 경우에는, 상기 강화학습 데이터 생성부(122)는 상기 강화학습 데이터를 해당 기계학습 네트워크에 적합한 포맷으로 변화하여 상기 학습부(200)로 제공할 수 있다.
- [0150] 또한 상기 학습모델 업데이트부(123)는, 상기 학습부(200)에서 강화학습을 수행하여, 상기 디바이스 제어용 학습네트워크에 대한 학습 파라미터를 제공 받아, 상기 제공받은 학습 파라미터를 상기 메모리(124)에 저장한 학습모델에 적용함으로써, 상기 학습모델을 업데이트한다.
- [0151] 즉, 상기 강화학습 에이전트(120)는, 미리 설정한 에피소드에 따라 상기 디바이스(300)를 제어하는 과정과, 상기 에피소드에 따라 강화학습 데이터를 생성하여 상기 학습부(200)로 제공하는 과정 및 상기 학습부(200)로부터 강화학습에 대한 학습 파라미터를 제공받아 상기 학습모델을 업데이트하는 과정을 반복하여 수행함으로써, 상기 디바이스(300)의 환경에 적응하여 상기 디바이스(300)를 정밀하게 제어할 수 있도록 한다.
- [0152] 도 7은 본 발명의 일 실시예에 따른 강화학습 기반 디바이스 원격 제어 시스템의 학습부에 대한 구성을 나타낸 블록도이다.
- [0153] 도 7에 도시한 바와 같이, 본 발명의 일 실시예에 따른 강화학습 기반 디바이스 원격 제어 시스템(10)의 학습부(200)는, 사용자 혹은 관리자에 의해 사전에 생성된 학습데이터를 학습하여 학습모델을 생성하고, 상기 생성한 학습모델을 상기 디바이스 원격 제어 장치(100)의 강화학습 에이전트(120)로 제공함으로써, 상기 학습모델을 통해 물리 환경에 위치하는 디바이스(300)를 원격에서 제어하도록 한다.
- [0154] 또한 상기 학습부(200)는, 에피소드가 종료될 때 마다 생성되는 강화학습 데이터를 상기 디바이스 원격 제어 장치(100)의 강화학습 에이전트(120)로부터 수신하는 강화학습 데이터 수신부(210), 상기 수신한 강화학습 데이터를 이용하여 상기 학습모델을 생성한 디바이스 제어용 학습네트워크에 대한 강화학습을 수행하는 강화 학습부(220) 및 상기 수행한 강화학습 결과에 대한 학습 파라미터를 상기 강화학습 에이전트(120)로 제공하는 학습 파라미터 제공부(230)를 포함하여 구성된다.
- [0155] 상기 강화학습 데이터는, 미리 설정한 에피소드 동안 상기 디바이스(300)를 제어한 결과를 누적하여 생성되는 것으로, 상기 강화학습 데이터 수신부(210)는 상기 에피소드가 종료될 때마다 상기 강화학습 에이전트(120)로부터 상기 강화학습 데이터를 수신하는 기능을 수행한다.
- [0156] 또한 상기 강화 학습부(220)는, 상기 수신한 강화학습 데이터를 이용하여 상기 디바이스 제어용 학습네트워크에 대한 강화학습을 수행한다.
- [0157] 즉, 상기 강화 학습부(220)는, 에피소드 동안 누적된 강화학습 데이터를 상기 디바이스 제어용 학습네트워크에 적용하여, 상기 디바이스 제어용 학습네트워크를 지속적으로 업데이트함으로써, 상기 디바이스(300)가 위치하는 물리 환경에 적응하여 상기 디바이스(300)를 정확하고 정밀하게 제어할 수 있도록 한다.
- [0158] 또한 상기 학습 파라미터 제공부(230)는, 강화학습을 수행한 결과, 해당 디바이스 제어용 학습네트워크의 학습 파라미터를 상기 강화학습 에이전트(120)로 제공하여, 상기 강화학습 에이전트(120)의 메모리(124)에 저장된 상기 학습모델에 상기 학습 파라미터를 적용할 수 있도록 함으로써, 상기 학습모델을 업데이트할 수 있도록 한다.
- [0159] 즉, 상기 학습모델은, 강화학습을 수행한 결과인 학습 파라미터를 통해 점점 고도화되어 상기 디바이스(300)를 신속하고 정밀하게 제어할 수 있도록 한다.
- [0160] 도 8은 본 발명의 일 실시예에 따른 강화학습 기반 디바이스 원격 제어 방법에 대한 절차를 나타낸 흐름도이다.
- [0161] 도 8에 도시한 바와 같이, 본 발명의 일 실시예에 따른 강화학습 기반 디바이스 원격 제어 시스템(10)을 통해 물리 환경에 위치하는 디바이스(300)를 사이버 환경에서 원격으로 제어하는 과정은 우선, 상기 강화학습 기반 디바이스 원격 제어 시스템(10)은 미리 설정한 에피소드에 따라 상기 디바이스(300)를 원격에서 제어하고자 하는 경우(S110), 사이버 환경부(110)를 통해 상기 디바이스(300)를 초기화하고, 상기 디바이스(300)로부터 초기화한 디바이스 상태정보를 수신 받아 강화학습 에이전트(120)로 제공한다(S120).
- [0162] 상기 초기화는, 상기 사이버 환경부(110)에서 상기 디바이스(300)를 초기화하기 위한 제어정보를 생성하여, 상기 디바이스(300)로 전송함으로써, 수행된다.

- [0163] 다음으로, 상기 강화학습 에이전트(120)에서, 상기 사이버 환경부(110)에서 제공 받은 디바이스 상태정보를 학습모델에 입력하여, 상기 디바이스를 제어하기 위한 제어정보를 생성한다(S130).
- [0164] 상기 학습모델은, 해당 디바이스(300)의 상태정보를 토대로 다음의 상태로 전이시키기 위한 적어도 하나 이상의 제어정보를 출력한다.
- [0165] 이때, 상기 제어정보는, 해당 디바이스의 상태정보를 토대로 상기 에피소드가 종료될 때 까지 모든 상태에 대한 리워드 정보의 총합이 높은 순으로 적어도 하나 이상으로 출력되며, 상기 강화학습 에이전트(120)는, 상기 리워드 정보의 총합이 제일 높은 제어정보를 선택함으로써, 상기 제어정보를 생성하게 된다.
- [0166] 한편 상기 학습모델은, 상기 강화학습 기반 디바이스 원격 제어 시스템(10)의 학습부(200)에서 사용자 혹은 관리자에 의해 생성된 학습데이터를 학습하여 생성되어, 상기 강화학습 에이전트(120)로 제공된 것이다.
- [0167] 다음으로, 상기 강화학습 에이전트(120)에서, 상기 생성한 제어정보를 사이버 환경부(110)로 제공하고, 상기 사이버 환경부(110)는, 상기 제공받은 제어정보를 상기 디바이스(300)로 전송함으로써, 상기 디바이스(300)를 원격에서 제어하는 기능을 수행한다(S140).
- [0168] 한편 상기 디바이스 상태정보 및 상기 제어정보는, 특정 통신방법을 통해 송수신되며, 네트워크 컨트롤러(400)에 의해 실시간으로 송수신되도록 제어되는 상술한 바와 같다.
- [0169] 다음으로, 상기 사이버 환경부(110)에서, 상기 디바이스(300)로부터 상기 제어정보를 토대로 동작되어 전이된 디바이스 상태정보를 수신하고(S150), 상기 수신한 디바이스 상태정보에 대한 리워드 정보를 생성하여, 상기 수신한 디바이스 상태정보 및 상기 생성한 해당 디바이스 상태정보에 대한 리워드 정보를 상기 강화학습 에이전트(120)로 제공한다(S160).
- [0170] 상기 리워드 정보는, 미리 설정한 임계범위를 기준으로 생성되는 것으로, 상기 제어정보를 토대로 상기 디바이스(300)가 제어되어 특정 상태로 전이된 결과(즉, 디바이스 상태정보)가 미리 설정한 임계범위 내인 경우, 플러스 보상값을 상기 디바이스 상태정보에 대해 설정하거나, 미리 설정한 임계범위를 초과하는 경우 마이너스 보상값을 상기 디바이스 상태정보에 대해 설정함으로써, 생성된다.
- [0171] 다음으로, 상기 강화학습 에이전트(120)에서, 상기 제공받은 디바이스 상태정보 및 리워드 정보를 토대로 강화학습 데이터를 생성한다(S170).
- [0172] 상기 강화학습 데이터는, 상기 제어정보를 수신 받은 상기 디바이스(300)에서 상기 제어정보에 따라 제어된 상태의 디바이스 상태정보를 수신하고, 이에 대한 리워드 정보를 생성할 때마다 생성된다.
- [0173] 즉, 상기 강화학습 데이터는, 상기 미리 설정한 에피소드가 종료될 때까지 누적되어 생성된다.
- [0174] 또한 상기 강화학습 데이터는, 상기 디바이스(300)의 현재 상태정보, 제어정보, 리워드 정보 및 다음 상태정보를 포함하는 상술한 바와 같다.
- [0175] 다음으로, 상기 미리 설정한 에피소드가 종료되지 않은 경우(S180), 상기 S130 단계 내지 S170단계를 반복하여 상기 디바이스(300)를 상기 에피소드에 따라 원격에서 제어할 수 있도록 한다.
- [0176] 한편, 상기 미리 설정한 에피소드가 종료된 경우(S180), 상기 강화학습 에이전트(120)는, 상기 생성한 강화학습 데이터를 학습부(200)로 제공하여, 상기 학습부(200)에서 강화학습을 수행할 수 있도록 한다(S190).
- [0177] 즉, 상기 학습부(200)는, 상기 제공받은 강화학습 데이터를 이용하여, 디바이스 제어용 학습네트워크에 대한 강화학습을 수행하며, 상기 강화학습을 수행한 결과에 대한 학습 파라미터를 상기 강화학습 에이전트(120)로 제공함으로써, 상기 학습 파라미터를 학습모델에 적용할 수 있도록 하여, 해당 학습모델을 고도화할 수 있도록 한다.
- [0178] 이상에서 설명한 바와 같이, 본 발명의 일 실시예에 따른 강화학습 기반 디바이스 원격 제어 시스템은, 물리 환경에 위치하는 디바이스와 사이버 환경에 위치하는 디바이스 원격 제어 장치 간의 상호 통신을 통해 상기 디바이스를 제어하기 위한 학습모델에 대한 지속적인 강화학습을 통해 상기 물리 환경에 적응하여 상기 디바이스를 정밀하고 효율적으로 제어할 수 있도록 하고, 상기 디바이스와 디바이스 원격 제어 장치 간의 원격 제어를 위한 데이터는 네트워크 컨트롤러를 통해 실시간으로 송수신 되도록 제어함으로써, 상기 디바이스를 원격에서 실시간으로 제어할 수 있도록 하는 효과가 있다.
- [0179] 상기에서는 본 발명에 따른 바람직한 실시예를 위주로 상술하였으나, 본 발명의 기술적 사상은 이에 한정되는

것은 아니며 본 발명의 각 구성요소는 동일한 목적 및 효과의 달성을 위하여 본 발명의 기술적 범위 내에서 변경 또는 수정될 수 있을 것이다.

[0180] 또한, 이상에서는 본 발명의 바람직한 실시예에 대하여 도시하고 설명하였지만, 본 발명은 상술한 특정의 실시예에 한정되지 아니하며, 청구범위에서 청구하는 본 발명의 요지를 벗어남이 없이 당해 발명이 속하는 기술분야에서 통상의 지식을 가진 자에 의해 다양한 변형 실시가 가능한 것은 물론이고, 이러한 변형 실시들은 본 발명의 기술적 사상이나 전망으로부터 개별적으로 이해되어서는 안 될 것이다.

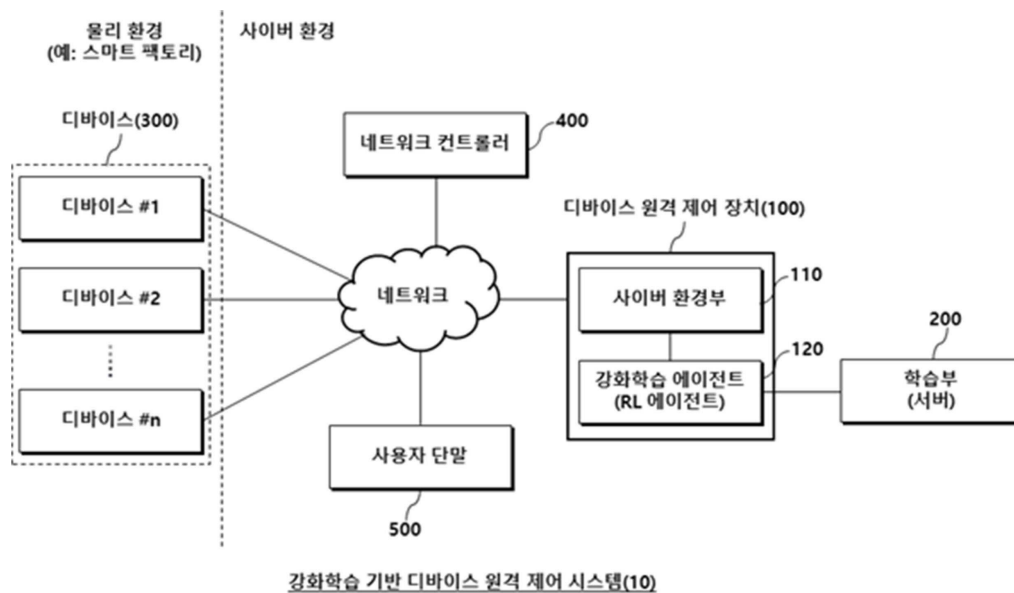
부호의 설명

[0181] 10: 강화학습 기반 디바이스 원격 제어 시스템

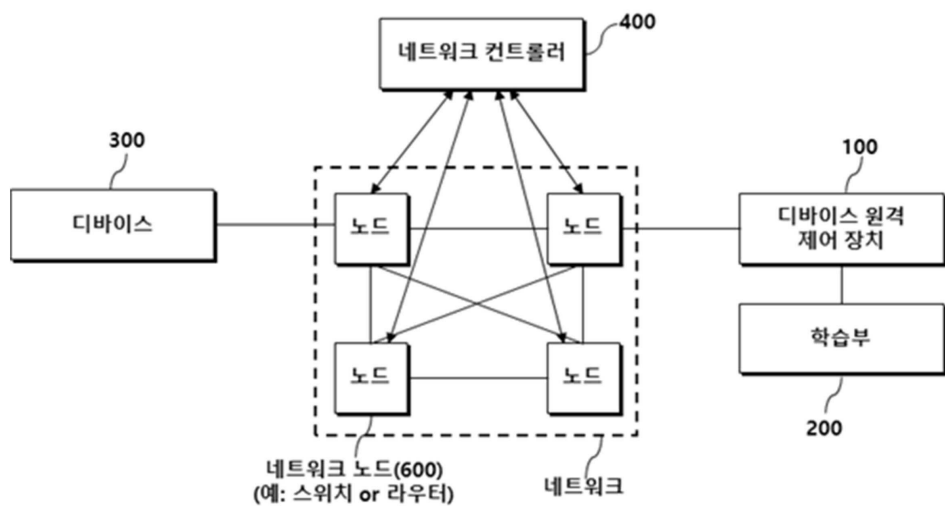
100: 디바이스 원격 제어 장치	110: 사이버 환경부
111: 제어정보 전송부	112: 디바이스 상태정보 수신부
113: 리워드 정보 생성부	114: 통신부
120: 강화학습 에이전트	121: 제어정보 수신부
122: 강화학습 데이터 생성부	123: 학습모델 업데이트부
124: 메모리	200: 학습부
210: 강화학습 데이터 수신부	220: 강화 학습부
230: 학습 파라미터 제공부	300: 디바이스
400: 네트워크 컨트롤러	500: 사용자 단말

도면

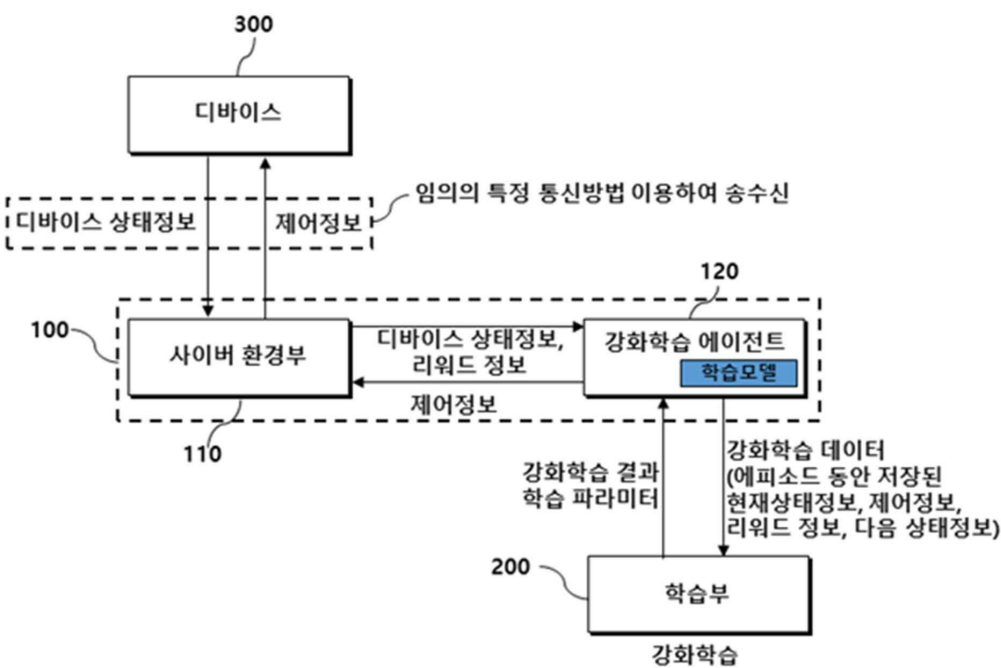
도면1



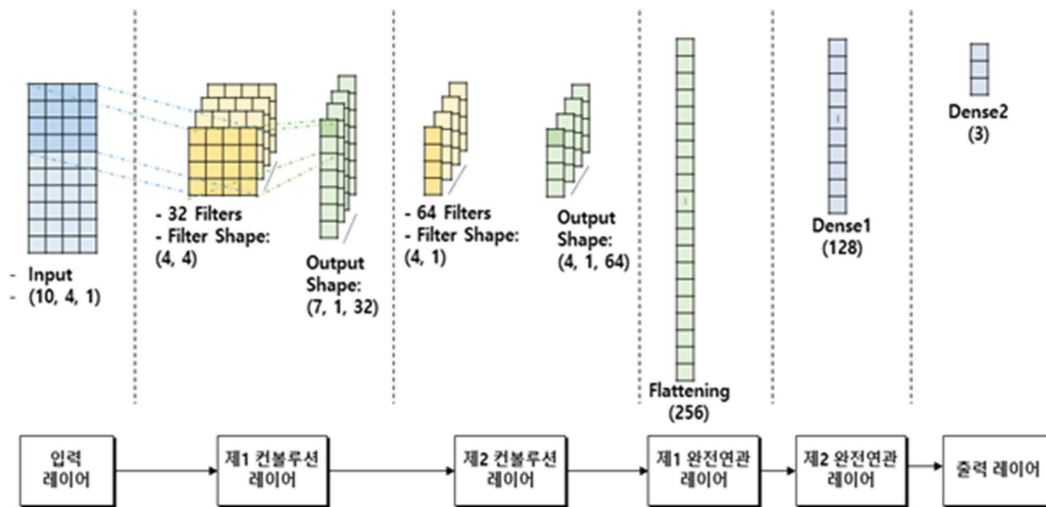
도면2



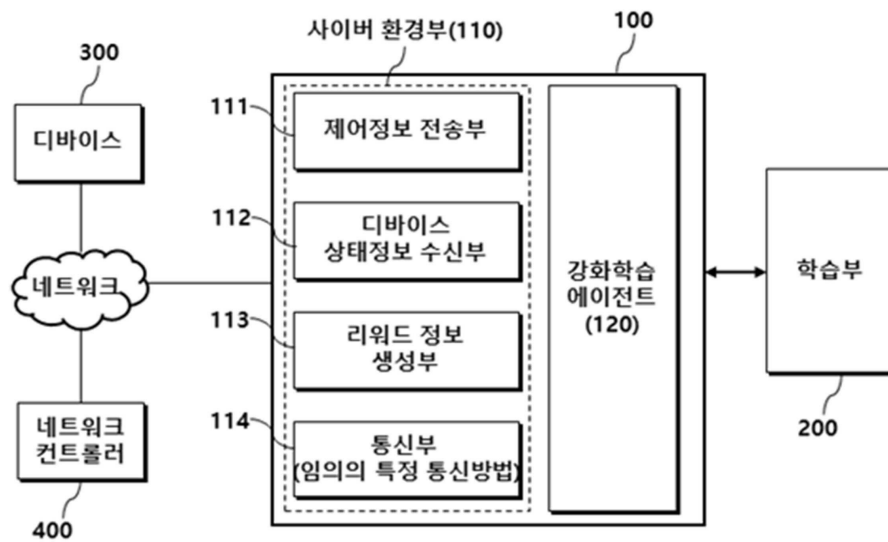
도면3



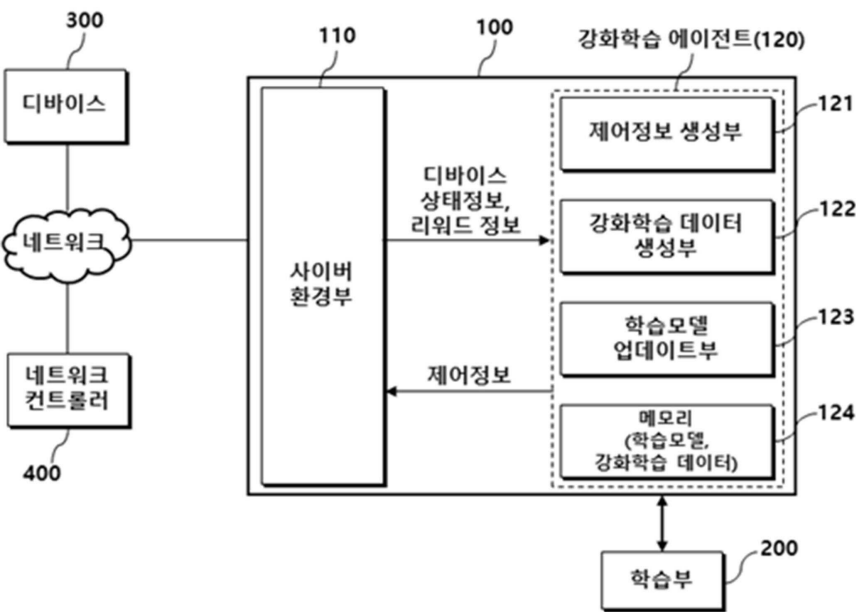
도면4



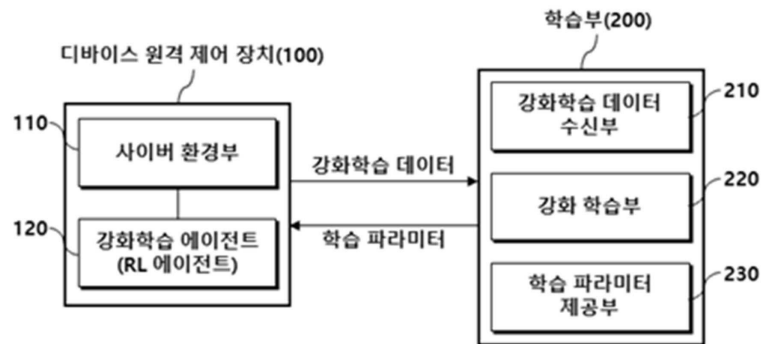
도면5



도면6



도면7



도면8

