



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2024년08월16일  
(11) 등록번호 10-2694987  
(24) 등록일자 2024년08월08일

(51) 국제특허분류(Int. Cl.)  
B25J 9/16 (2006.01) B25J 13/08 (2006.01)  
B25J 19/02 (2006.01) G06T 17/00 (2006.01)  
G06T 3/40 (2024.01) G06T 7/33 (2017.01)  
G06T 7/70 (2017.01)  
(52) CPC특허분류  
B25J 9/1697 (2013.01)  
B25J 13/08 (2013.01)  
(21) 출원번호 10-2023-0187604  
(22) 출원일자 2023년12월20일  
심사청구일자 2023년12월20일  
(56) 선행기술조사문헌  
JP2021163503 A\*  
JP2022504009 A\*  
KR1020180129301 A\*  
KR1020200119930 A\*  
\*는 심사관에 의하여 인용된 문헌

(73) 특허권자  
주식회사 플라잇  
경기도 성남시 분당구 대왕판교로 660, 유스페이스1 B동 901호 (삼평동)  
(72) 발명자  
최재우  
경기도 성남시 분당구 장미로 101, 809동 401호  
한가영  
경기도 광주시 오포읍 새말길 19-20, 5동 403호  
(74) 대리인  
특허법인엠에이피에스

전체 청구항 수 : 총 8 항

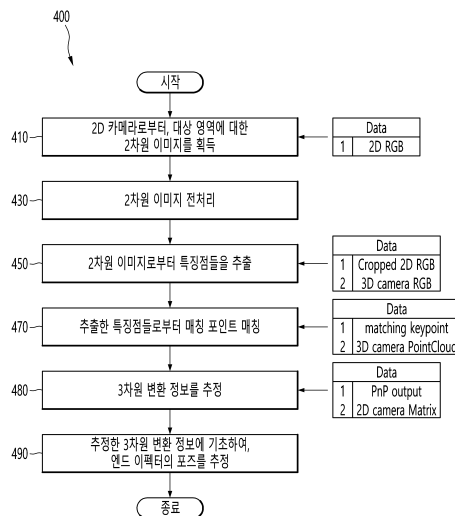
심사관 : 박성용

(54) 발명의 명칭 인공지능을 이용하여 객체에 대한 로봇의 엔드 이펙터의 포즈를 추정하는 포즈 추정 장치 및 방법과 이를 포함하는 로봇 제어 시스템

(57) 요약

본 개시의 일 실시예에 따른 인공지능을 이용하여 객체에 대한 로봇의 엔드 이펙터의 포즈를 추정하는 포즈 추정 장치는, 적어도 하나의 프로세서, 및 컴퓨터 프로그램 코드를 포함하는 적어도 하나의 메모리를 포함한다. 상기 적어도 하나의 메모리와 상기 컴퓨터 프로그램 코드는 상기 적어도 하나의 프로세서를 통해 상기 포즈 추정 장치가, 2D 카메라로부터, 대상 영역에 대응하는 2차원 이미지를 획득하고, 기학습된 인공지능 모델에 기초하여, 상기 획득한 2차원 이미지에 대응하는 3차원 변환 정보를 추정하고, 상기 추정한 3차원 변환 정보에 기초하여, 상기 엔드 이펙터의 포즈를 추정하도록 구성된다. 상기 인공지능 모델은, 상기 2차원 이미지를 기저장된 3차원 비교 이미지에 매칭함으로써 상기 3차원 변환 정보를 추정하도록 학습된 것이다.

대표도 - 도4



(52) CPC특허분류

*B25J 19/023* (2013.01)

*B25J 9/161* (2013.01)

*B25J 9/163* (2013.01)

*B25J 9/1653* (2013.01)

*B25J 9/1664* (2013.01)

*G06T 17/00* (2013.01)

*G06T 3/40* (2024.01)

*G06T 7/33* (2017.01)

*G06T 7/70* (2017.01)

---

## 명세서

### 청구범위

#### 청구항 1

인공지능을 이용하여 객체에 대한 로봇의 엔드 이펙터의 포즈를 추정하는 포즈 추정 장치에 있어서,

적어도 하나의 프로세서; 및

컴퓨터 프로그램 코드를 포함하는 적어도 하나의 메모리를 포함하고,

상기 적어도 하나의 메모리와 상기 컴퓨터 프로그램 코드는 상기 적어도 하나의 프로세서를 통해 상기 포즈 추정 장치가:

2D 카메라로부터, 대상 영역에 대응하는 2차원 이미지를 획득하고,

기학습된 인공지능 모델에 기초하여, 상기 획득한 2차원 이미지에 대응하는 3차원 변환 정보를 추정하고,

기저장된 3차원 비교 이미지에 대응하는 3D 카메라와 상기 엔드 이펙터 사이의 상대적 위치에 대응하는 제1 변환 매트릭스 및 상기 2차원 이미지에 대응하는 상기 2D 카메라와 상기 로봇의 베이스 사이의 상대적 위치에 대응하는 제2 변환 매트릭스에 기초하여, 상기 추정한 3차원 변환 정보로부터 상기 엔드 이펙터의 포즈를 추정하도록 구성되고,

상기 인공지능 모델은, 상기 2차원 이미지를 상기 기저장된 3차원 비교 이미지에 매칭함으로써 상기 3차원 변환 정보를 추정하도록 학습된 것인,

포즈 추정 장치.

#### 청구항 2

제 1 항에 있어서,

상기 적어도 하나의 메모리와 상기 컴퓨터 프로그램 코드는 상기 적어도 하나의 프로세서를 통해 상기 포즈 추정 장치가:

상기 2차원 이미지로부터 일부 영역을 추출하고, 상기 추출한 일부 영역을 그레이 스케일로 전처리하도록 구성된 것인,

포즈 추정 장치.

#### 청구항 3

삭제

#### 청구항 4

제 1 항에 있어서,

상기 인공지능 모델은, 상기 2차원 이미지로부터 특징점들을 추출하고, 상기 추출한 특징점들에 기초하여 상기 2차원 이미지를 상기 기저장된 3차원 비교 이미지에 매칭하도록 학습된 것인,

포즈 추정 장치.

#### 청구항 5

제 4 항에 있어서,

상기 인공지능 모델은, 그래프 신경망에 기초하여, 상기 추출한 특징점들로부터 상기 2차원 이미지와 상기 기저장된 3차원 비교 이미지를 매칭하는 매칭 포인트들을 포함하는 상기 3차원 변환 정보를 추정하도록 학습된 것인,

포즈 추정 장치.

#### 청구항 6

제 5 항에 있어서,

상기 적어도 하나의 메모리와 상기 컴퓨터 프로그램 코드는 상기 적어도 하나의 프로세서를 통해 상기 포즈 추정 장치가:

상기 매칭 포인트들에 기초하여, PnP(Perspective N Points) 방식을 사용함으로써 상기 기저장된 3차원 비교 이미지에 대응하는 3D 카메라에 대비한 상기 2차원 이미지에 대응하는 상기 2D 카메라의 자세 데이터를 추정하도록 구성된 것인,

포즈 추정 장치.

#### 청구항 7

제 6 항에 있어서,

상기 적어도 하나의 메모리와 상기 컴퓨터 프로그램 코드는 상기 적어도 하나의 프로세서를 통해 상기 포즈 추정 장치가:

RANSAC(random sample consensus) 기법을 이용함으로써, 상기 매칭 포인트들에 기초하여 적어도 하나 이상의 모델을 추정하고, 상기 추정한 모델 중 최적 모델을 선택하도록 구성된 것인,

포즈 추정 장치.

#### 청구항 8

인공지능을 이용하여 객체에 대한 로봇의 엔드 이펙터의 포즈를 추정하는 포즈 추정 방법에 있어서,

2D 카메라로부터, 대상 영역에 대한 2차원 이미지를 획득하는 동작;

기학습된 인공지능 모델에 기초하여, 상기 2차원 이미지에 대응하는 3차원 변환 정보를 추정하는 동작; 및

기저장된 3차원 비교 이미지에 대응하는 3D 카메라와 상기 엔드 이펙터 사이의 상대적 위치에 대응하는 제1 변환 매트릭스 및 상기 2차원 이미지에 대응하는 상기 2D 카메라와 상기 로봇의 베이스 사이의 상대적 위치에 대응하는 제2 변환 매트릭스에 기초하여, 상기 추정한 3차원 변환 정보로부터 상기 엔드 이펙터의 포즈를 추정하는 동작을 포함하고,

상기 인공지능 모델은, 상기 2차원 이미지를 상기 기저장된 3차원 비교 이미지에 매칭함으로써 상기 3차원 변환 정보를 추정하도록 학습된 것인,

포즈 추정 방법.

#### 청구항 9

인공지능을 이용하여 객체에 대한 로봇의 엔드 이펙터의 포즈를 추정하는 포즈 추정 장치를 포함하는 로봇 제어 시스템에 있어서,

대상 영역에 대한 2차원 이미지를 획득하도록 구성된 2D 카메라;

상기 2D 카메라로부터 상기 2차원 이미지를 획득하고, 기학습된 인공지능 모델에 기초하여 상기 획득한 2차원 이미지에 대응하는 3차원 변환 정보를 추정하고, 기저장된 3차원 비교 이미지에 대응하는 3D 카메라와 상기 엔드 이펙터 사이의 상대적 위치에 대응하는 제1 변환 매트릭스 및 상기 2차원 이미지에 대응하는 상기 2D 카메라와 상기 로봇의 베이스 사이의 상대적 위치에 대응하는 제2 변환 매트릭스에 기초하여 상기 추정한 3차원 변환 정보로부터 상기 엔드 이펙터의 포즈를 추정하도록 구성된 포즈 추정 장치; 및

베이스 및 회전 또는 이동 가능한 엔드 이펙터가 구비되고, 상기 추정한 3차원 변환 정보에 기초하여 상기 엔드 이펙터를 제어하도록 구성된 로봇 장치를 포함하고,

상기 인공지능 모델은, 상기 2차원 이미지를 상기 기저장된 3차원 비교 이미지에 매칭함으로써 상기 3차원 변환 정보를 추정하도록 학습된 것인,

로봇 제어 시스템.

## 발명의 설명

### 기술 분야

[0001] 본 개시는, 인공지능을 이용하여 객체에 대한 로봇의 엔드 이펙터의 포즈를 추정하는 포즈 추정 장치 및 방법과 이를 포함하는 로봇 제어 시스템에 관한 것이다.

### 배경 기술

[0002] 로봇은 제품 생산에서부터 출하까지의 공정 내 작업을 수행하는 주요 장비로서, 사람보다 더 큰 공정 효율을 나타내어 폭넓은 산업 기술 분야에서 사용되고 있다. 특히, 제조용 로봇은 작업자가 손으로 파지하거나 공구로 들어 옮기는 작업에 따른 안전 사고를 예방하고, 작업의 효율을 향상시킬 수 있다.

[0003] 이러한 로봇은 다양하고 정밀한 작업을 위하여 다관절로 구성되고, 이에 따라 엔드 이펙터는 복수 개의 축 방향에 따른 이동 및 회전이 가능하다. 예를 들어, 엔드 이펙터는 X축, Y축 및 Z축에 따른 이동 및 회전이 가능한 6 자유도(DoF)를 가질 수 있다.

[0004] 로봇의 동작을 제어하기 위해서는, 엔드 이펙터의 포즈에 관한 정보가 요구되고, 이를 위하여 종래에는 3D 카메라를 이용함으로써 자유도에 따른 엔드 이펙터의 포즈를 감지하였다. 다만, 상대적으로 값 비싼 3D 카메라를 이용함에 따라, 로봇 시스템의 적용에 따른 단가가 높아지는 문제가 있었다.

## 발명의 내용

### 해결하려는 과제

[0005] 본 개시의 일 목적은, 기학습된 인공지능 모델에 기초하여, 2D 카메라를 이용하여 획득한 2차원 이미지로부터 엔드 이펙터의 포즈를 추정하기 위함이다.

[0006] 다만, 본 실시예가 이루고자 하는 기술적 과제는 상기된 바와 같은 기술적 과제들로 한정되지 않으며, 또 다른 기술적 과제들이 존재할 수 있다.

### 과제의 해결 수단

[0007] 본 개시의 일 실시예에 따른 인공지능을 이용하여 객체에 대한 로봇의 엔드 이펙터의 포즈를 추정하는 포즈 추정 장치는, 적어도 하나의 프로세서, 및 컴퓨터 프로그램 코드를 포함하는 적어도 하나의 메모리를 포함할 수 있다. 상기 적어도 하나의 메모리와 상기 컴퓨터 프로그램 코드는 상기 적어도 하나의 프로세서를 통해 상기 포즈 추정 장치가, 2D 카메라로부터, 대상 영역에 대응하는 2차원 이미지를 획득하고, 기학습된 인공지능 모델에 기초하여, 상기 획득한 2차원 이미지에 대응하는 3차원 변환 정보를 추정하고, 상기 추정한 3차원 변환 정보에 기초하여, 상기 엔드 이펙터의 포즈를 추정하도록 구성될 수 있다. 상기 인공지능 모델은, 상기 2차원 이미지를 기저장된 3차원 비교 이미지에 매칭함으로써 상기 3차원 변환 정보를 추정하도록 학습된 것일 수 있다.

[0008] 본 개시의 일 실시예에 따른 인공지능을 이용하여 객체에 대한 로봇의 엔드 이펙터의 포즈를 추정하는 포즈 추정 방법은, 인공지능을 이용하여 객체에 대한 로봇의 엔드 이펙터의 포즈를 추정하는 포즈 추정 방법에 있어서, 2D 카메라로부터, 대상 영역에 대한 2차원 이미지를 획득하는 동작, 기학습된 인공지능 모델에 기초하여, 상기 2차원 이미지에 대응하는 3차원 변환 정보를 추정하는 동작, 및 상기 추정된 3차원 변환 정보에 기초하여, 상기 엔드 이펙터의 포즈를 추정하는 동작을 포함할 수 있다. 상기 인공지능 모델은, 상기 2차원 이미지를 기저장된 3차원 비교 이미지에 매칭함으로써 상기 3차원 변환 정보를 추정하도록 학습된 것일 수 있다.

[0009] 본 개시의 일 실시예에 따른 인공지능을 이용하여 객체에 대한 로봇의 엔드 이펙터의 포즈를 추정하는 포즈 추정 장치를 포함하는 로봇 제어 시스템은, 대상 영역에 대한 2차원 이미지를 획득하도록 구성된 2D 카메라, 상기 2D 카메라로부터 상기 2차원 이미지를 획득하고, 기학습된 인공지능 모델에 기초하여 상기 획득한 2차원 이미지에 대응하는 3차원 변환 정보를 추정하고, 상기 추정된 3차원 변환 정보에 기초하여 상기 엔드 이펙터의 포즈를 추정하도록 구성된 포즈 추정 장치, 및 베이스 및 회전 또는 이동 가능한 엔드 이펙터가 구비되고, 상기 추정된 3차원 변환 정보에 기초하여 상기 엔드 이펙터를 제어하도록 구성된 로봇 장치를 포함할 수 있다. 상기 인공지능 모델은, 상기 2차원 이미지를 기저장된 3차원 비교 이미지에 매칭함으로써 상기 3차원 변환 정보를 추정하도록 학습된 것일 수 있다.

**발명의 효과**

[0010] 본 개시의 일 실시예에 따르면, 인공지능을 이용한 학습 모델에 기초하여 2차원 이미지에 대응하는 3차원 변환 정보를 추정함으로써, 3차원 공간의 깊이 데이터를 획득하고, 이에 따라 엔드 이펙터의 포즈를 추정할 수 있다.

[0011] 또한, 본 개시의 일 실시예에 따르면 상대적으로 저렴한 2D 카메라를 이용하여 로봇의 엔드 이펙터의 포즈를 추정함으로써, 로봇 시스템의 단가를 낮출 수 있고, 대규모 데이터 처리에 소모되는 컴퓨팅 리소스를 절감할 수 있다.

**도면의 간단한 설명**

- [0012] 도 1은, 본 개시의 일 실시예에 따른, 포즈 추정 장치의 구성도를 도시한 것이다.
- 도 2는, 본 개시의 일 실시예에 따른, 포즈 추정 장치를 포함하는 로봇 제어 시스템의 구성도를 도시한 것이다.
- 도 3은, 본 개시의 일 실시예에 따른 로봇 장치의 예시를 도시한 것이다.
- 도 4는, 본 개시의 일 실시예에 따른 포즈 추정 방법의 흐름도이다.
- 도 5는, 본 개시의 일 실시예에 따른 2차원 이미지의 예시이다.
- 도 6은, 본 개시의 일 실시예에 따른 인공지능 모델의 예시이다.
- 도 7은, 본 개시의 일 실시예에 따른 2차원 이미지 및 기저장된 3차원 비교 이미지 사이의 매칭된 특징점들의 예시이다.

**발명을 실시하기 위한 구체적인 내용**

[0013] 아래에서는 첨부한 도면을 참조하여 본원이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 본원의 실시예를 상세히 설명한다. 그러나 본원은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본원을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

[0014] 본원 명세서 전체에서, 어떤 부재가 다른 부재 "상에" 위치하고 있다고 할 때, 이는 어떤 부재가 다른 부재에 접해 있는 경우 뿐 아니라 두 부재 사이에 또 다른 부재가 존재하는 경우도 포함한다.

[0015] 본원 명세서 전체에서, 어떤 부분이 어떤 구성 요소를 "포함" 한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성 요소를 제외하는 것이 아니라 다른 구성 요소를 더 포함할 수 있는 것을 의미한다.

[0016] 본원 명세서 전체에서 사용하는 정도의 용어 "약", "실질적으로" 등은 언급된 의미에 고유한 제조 및 물질 허용 오차가 제시될 때 그 수치에서 또는 그 수치에 근접한 의미로 사용되고, 본원의 이해를 돕기 위해 정확하거나 절대적인 수치가 언급된 개시 내용을 비양심적인 침해자가 부당하게 이용하는 것을 방지하기 위해 사용된다. 본원 명세서 전체에서 사용하는 정도의 용어 "~(하는) 단계" 또는 "~의 단계"는 "~를 위한 단계"를 의미하지 않

는다.

- [0017] 본원 명세서 전체에서, 마쿠시 형식의 표현에 포함된 "이들의 조합(들)"의 용어는 마쿠시 형식의 표현에 기재된 구성 요소들로 이루어진 군에서 선택되는 하나 이상의 혼합 또는 조합을 의미하는 것으로서, 상기 구성 요소들로 이루어진 군에서 선택되는 하나 이상을 포함하는 것을 의미한다.
- [0018] 본원 명세서 전체에서, "A 및/또는 B"의 기재는 "A 또는 B, 또는 A 및 B"를 의미한다.
- [0019] 이하, 첨부된 도면을 참조하여 본원의 구현에 및 실시예를 상세히 설명한다. 그러나, 본원이 이러한 구현에 및 실시예와 도면에 제한되지 않을 수 있다.
- [0020] 도 1은, 본 개시의 일 실시예에 따른, 포즈 추정 장치의 구성도를 도시한 것이다. 도 2는, 본 개시의 일 실시예에 따른, 포즈 추정 장치(100)를 포함하는 로봇 제어 시스템(200)의 구성도를 도시한 것이다.
- [0021] 도 1 내지 도 2를 참조하면, 일 실시예에 따른, 포즈 추정 장치(100)를 포함하는 로봇 제어 시스템(200)은 포즈 추정 장치(100), 2D 카메라(210) 및/또는 로봇 장치(230)를 포함할 수 있다.
- [0022] 일 실시예에 따른 포즈 추정 장치(100)는, 프로세서(110) 및/또는 메모리(120)를 포함할 수 있다.
- [0023] 프로세서(110)는, 예를 들어 소프트웨어(예: 프로그램)를 실행하여 프로세서(110)에 연결된 포즈 추정 장치(100)의 적어도 하나의 다른 구성요소(예: 하드웨어 또는 소프트웨어 구성요소)를 제어할 수 있고, 다양한 데이터 처리 또는 연산을 수행할 수 있다. 일 실시예에 따르면, 데이터 처리 또는 연산의 적어도 일부로서, 프로세서(110)는 다른 구성요소로부터 수신된 명령 또는 데이터를 휘발성 메모리에 저장하고, 휘발성 메모리에 저장된 명령 또는 데이터를 처리하고, 결과 데이터를 비휘발성 메모리에 저장할 수 있다.
- [0024] 일 실시예에 따르면, 프로세서(110)는 메인 프로세서(예: 중앙 처리 장치 또는 어플리케이션 프로세서) 또는 이와는 독립적으로 또는 함께 운영 가능한 보조 프로세서(예: 그래픽 처리 장치, 신경망 처리 장치(NPU: neural processing unit), 이미지 시그널 프로세서, 센서 허브 프로세서, 또는 커뮤니케이션 프로세서)를 포함할 수 있다. 예를 들어, 보조 프로세서는 메인 프로세서보다 저전력을 사용하거나, 지정된 기능에 특화되도록 설정될 수 있고, 보조 프로세서는 메인 프로세서와 별개로, 또는 그 일부로서 구현될 수 있다.
- [0025] 보조 프로세서는, 예를 들어 메인 프로세서가 인액티브(예: 슬립) 상태에 있는 동안 메인 프로세서를 대신하여, 또는 메인 프로세서가 액티브(예: 어플리케이션 실행) 상태에 있는 동안 메인 프로세서와 함께, 포즈 추정 장치(100)의 구성요소들 중 적어도 하나의 구성요소와 관련된 기능 또는 상태들의 적어도 일부를 제어할 수 있다.
- [0026] 일 실시예에 따르면, 보조 프로세서(예: 이미지 시그널 프로세서 또는 커뮤니케이션 프로세서)는 기능적으로 관련 있는 다른 구성요소의 일부로서 구현될 수 있다. 일 실시예에 따르면, 보조 프로세서(예: 신경망 처리 장치)는 인공지능 모델의 처리에 특화된 하드웨어 구조를 포함할 수 있다.
- [0027] 인공지능 모델은 기계 학습을 통해 생성될 수 있다. 이러한 학습은, 예를 들어, 인공지능 모델이 수행되는 포즈 추정 장치(100) 자체에서 수행될 수 있고, 별도의 서버를 통해 수행될 수도 있다. 학습 알고리즘은, 예를 들어, 지도형 학습(supervised learning), 비지도형 학습(unsupervised learning), 준지도형 학습(semi-supervised learning) 또는 강화 학습(reinforcement learning)을 포함할 수 있으나, 전술한 예에 한정되지 않는다. 인공지능 모델은, 복수의 인공 신경망 레이어들을 포함할 수 있다. 인공 신경망은 심층 신경망(DNN: deep neural network), CNN(convolutional neural network), RNN(recurrent neural network), RBM(restricted boltzmann machine), DBN(deep belief network), BRDNN(bidirectional recurrent deep neural network), 심층 Q-네트워크(deep Q-networks) 또는 상기 중 둘 이상의 조합 중 하나일 수 있으나, 전술한 예에 한정되지 않는다. 인공지능 모델은 하드웨어 구조 이외에, 추가적으로 또는 대체적으로, 소프트웨어 구조를 포함할 수 있다.
- [0028] 메모리(120)는, 포즈 추정 장치(100)의 적어도 하나의 구성요소(예: 프로세서(110))에 의해 사용되는 다양한 데이터를 저장할 수 있다. 데이터는, 예를 들어, 소프트웨어(예: 프로그램) 및, 이와 관련된 명령에 대한 입력 데이터 또는 출력 데이터를 포함할 수 있다. 메모리(120)는, 휘발성 메모리 또는 비휘발성 메모리를 포함할 수 있다. 프로그램은 메모리(120)에 소프트웨어로서 저장될 수 있으며, 예를 들면, 운영 체제, 미들 웨어 또는 어플리케이션을 포함할 수 있다.
- [0029] 일 실시예에 따른 2D 카메라(210)는, 대상 영역에 대응하는 2차원 이미지를 획득하도록 구성될 수 있다. 예를 들어, 2차원 이미지는 평면 상의 위치 데이터를 갖는 이미지로, 깊이 데이터를 포함하지 않을 수 있다.
- [0030] 일 실시예에 따른 로봇 장치(230)는, 프로세서(231), 메모리(233), 베이스(235), 구동모듈(237) 및/또는 엔드



이펙터(239)를 포함할 수 있다. 일 실시예로, 로봇 장치(230)의 프로세서(231) 및/또는 메모리(233)는 별도의 제어 장치에 구비될 수도 있고, 로봇 장치(230) 내부에 일체화될 수 있다. 프로세서(231) 및/또는 메모리(233)에 관한 설명은 상기 포즈 추정 장치(100)에 관한 설명으로 대체한다.

- [0031] 도 3은, 본 개시의 일 실시예에 따른 로봇 장치(230)의 예시를 도시한 것이다.
- [0032] 도 3을 더 참조하면, 일 실시예에 따른 로봇 장치(230)는, 베이스(235), 구동모듈(237) 및/또는 엔드 이펙터(239)를 포함할 수 있다.
- [0033] 일 실시예로, 베이스(235)는 지면 또는 외부 장치에 고정적으로 결합될 수 있다. 일 실시예로, 베이스(235)의 좌표계(R)는 고정된 좌표계일 수 있다.
- [0034] 일 실시예로, 엔드 이펙터(239)는 구동모듈(237)을 통하여 이동 또는 회전 가능하도록 베이스(235)에 결합될 수 있다. 일 실시예로, 엔드 이펙터(239)의 좌표계(X)는 베이스(235)의 좌표계(R)에 대비하여 회전 또는 이동될 수 있다.
- [0035] 일 실시예로, 엔드 이펙터(239)는 로봇 장치(230)가 작업을 할 때, 작업 대상인 객체에 직접 작용하는 기능을 가진 부분일 수 있다. 예를 들어, 엔드 이펙터(239)에는 그리퍼, 용접 토치, 스프레이건, 너트 러너 등이 구비될 수 있다.
- [0036] 일 실시예로, 구동모듈(237)은 베이스(235)와 엔드 이펙터(239) 사이에 구비되어, 고정된 베이스(235)를 기준으로 엔드 이펙터(239)를 이동 또는 회전시킬 수 있다. 예를 들어, 구동모듈(237)은 전력에 의해 구동되어 엔드 이펙터(239)를 이동 또는 회전시키는 모터일 수 있다.
- [0037] 일 실시예로, 구동모듈(237)은 엔드 이펙터(239)를 복수의 방향으로 회전 또는 이동 가능하도록 복수 개로 구비될 수 있다. 예를 들어, 엔드 이펙터(239)는 X축, Y축 및 Z축에 따른 이동 및 회전이 가능한 6 자유도(DoF)를 가질 수 있다.
- [0038] 일 실시예로, 2D 카메라(210)는 베이스(235)와 마찬가지로 지면 또는 외부 장치에 고정적으로 결합될 수도 있으나, 엔드 이펙터(239)의 이동 또는 회전에 대응하여 이동 또는 회전될 수 있다. 2D 카메라(210)의 좌표계(C)는 고정될 수도 있고, 이동 또는 회전될 수도 있다.
- [0039] 일 실시예로, 2D 카메라(210)는 FOV(화각, Field of View)가 로봇 장치(230)가 위치하는 대상 영역을 포함하도록 배치될 수 있다. 일 실시예로, 2D 카메라(210)는 FOV가 로봇 장치(230)의 엔드 이펙터(239)가 위치 가능한 영역을 향하도록 배치될 수 있다. 일 실시예로, 2D 카메라(210)는 대상 영역에 대응하는 2차원 이미지를 실시간으로 획득하고, 포즈 추정 장치(100)로 전송할 수 있다.
- [0040] 일 실시예로, 엔드 이펙터(239)에는 2D 카메라(210)의 좌표계와 로봇의 좌표계를 캘리브레이션(calibration)하기 위한 마커(240)가 부착될 수 있다. 일 실시예로, 마커(240)는 2차원 이미지에 대응하여 후술하는 3차원 변환 정보를 획득하기 용이하도록 부착될 수도 있다. 예를 들어, 마커(240)는 아르코(Chess) 마커 또는 차르코(Charuco) 마커와 같은 체스보드 마커가 이용될 수 있으나, 이에 한정되는 것은 아니며 공지된 다양한 마커가 이용될 수 있다.
- [0041] 도 4는, 본 개시의 일 실시예에 따른 포즈 추정 방법의 흐름도(400)이다. 도 5는, 본 개시의 일 실시예에 따른 2차원 이미지(2D Image)의 예시이다.
- [0042] 도 4 내지 도 5를 참조하면, 일 실시예에 따른 포즈 추정 장치(100, 예: 프로세서(110))는, 동작 410에서, 2D 카메라(210)로부터, 대상 영역에 대한 2차원 이미지(2D Image, 예: 2D RGB)를 획득할 수 있다.
- [0043] 일 실시예로, 2D 카메라(210)로부터 획득한 2차원 이미지(2D Image)에는 로봇의 엔드 이펙터(239)의 이미지가 포함될 수 있다.
- [0044] 일 실시예에 따른 포즈 추정 장치(100)는, 동작 430에서, 2차원 이미지(2D Image)를 전처리할 수 있다.
- [0045] 일 실시예에 따른 포즈 추정 장치(100)는, 2D 카메라(210)로부터 획득한 2차원 이미지(2D Image)로부터 대상 영역에 위치한 로봇의 엔드 이펙터(239)에 대응하는 일부 영역을 추출하고, 추출한 일부 영역을 그레이 스케일로 전처리할 수 있다.
- [0046] 일 실시예로, 포즈 추정 장치(100)는 획득한 2차원 이미지(2D Image)에서 로봇의 엔드 이펙터(239)를 감지하고, 감지한 엔드 이펙터(239)를 포함하는 일부 영역(ROI: region of interest)을 절단(crop)하여 추출할 수 있다.



예를 들어, 일부 영역은 바운딩 박스(BBox)일 수 있다.

- [0047] 일 실시예로, 포즈 추정 장치(100)는 추출한 2차원 이미지(2D Image)의 일부 영역을 그레이 스케일(grey scale)로 처리할 수 있다. 이에 따라, 단일 채널을 갖는 그레이 스케일은 컬러 이미지에 대비하여 채널 수가 적어 계산을 효율적으로 수행할 수 있으며, 픽셀 값의 세기에 집중하여 특징을 학습 및 추론할 수 있어 간소화될 수 있다. 또한, 후술하는 포인트 클라우드의 매칭에서 조명 등의 환경 조건이 변화함에 따라 색상 데이터에 발생하는 차이를 감소시킬 수 있다. 또한, 그레이 스케일로 처리된 이미지는 구조적인 특징(예: 질감)을 강조하는데 효과적일 수 있다.
- [0048] 일 실시예에 따른 포즈 추정 장치(100)는, 동작 450에서, 인공지능 모델에 기초하여 2차원 이미지(예: Cropped 2D RGB)로부터 특징점들을 추출할 수 있다.
- [0049] 일 실시예로, 특징점은 이미지의 구석, 가장자리 및/또는 질감의 변화 지점일 수 있다.
- [0050] 일 실시예로, 포즈 추정 장치(100)는 2차원 이미지(예: Cropped 2D RGB)로부터 3차원 비교 이미지(3D camera RGB)에 가장 매칭되는 포인트를 특징점으로 추출할 수 있다. 일 실시예로, 3차원 비교 이미지(예: 3D camera RGB)는 별도의 장치에 포함된 3D 카메라(미도시)에 의해 획득된 것으로, 포즈 추정 장치(100, 예: 메모리(120))에 기저장될 수 있다.
- [0051] 일 실시예에 따른 포즈 추정 장치(100)는, 동작 470에서, 인공지능 모델에 기초하여 추출한 특징점들(예: matching keypoint)로부터 3차원 비교 이미지의 매칭 포인트(예: 3D camera PointCloud)를 매칭할 수 있다.
- [0052] 도 6은, 본 개시의 일 실시예에 따른 인공지능 모델의 예시이다. 도 7은, 본 개시의 일 실시예에 따른 2차원 이미지(2D Image) 및 기저장된 3차원 비교 이미지(3D Image) 사이의 매칭된 특징점들의 예시이다.
- [0053] 도 6 및 도 7을 더 참조하면, 일 실시예로, 인공지능 모델은, 2차원 이미지(2D Image)로부터 특징점들을 추출하고, 추출한 특징점들에 기초하여 2차원 이미지(2D Image)를 기저장된 3차원 비교 이미지(3D Image)에 매칭하도록 학습된 것일 수 있다.
- [0054] 일 실시예로, 인공지능 모델은, 그래프 신경망(Graph Neural Network)을 활용하여 이미지 간의 특징점을 매칭하는데 최적화된 딥러닝 모델일 수 있다. 일 실시예로, 인공지능 모델은 각각의 특징점을 그래프의 노드로 간주하고, 이들 간의 상호작용을 엣지로 표현할 수 있다. 인공지능 모델은 이러한 과정을 통해 복잡한 패턴과 관계를 효율적으로 학습하고, 더 정교한 매칭 결과를 도출할 수 있다.
- [0055] 예를 들어, 도 6에 도시한 LightGlue 아키텍처에 따르면, 이미지들(images)로부터 한 쌍의 입력 로컬 특징(local features,  $d, p$ )이 주어지면, 각 레이어는 위치 인코딩을 사용하여 셀프-어텐션 또는 크로스-어텐션을 기반으로 하는 컨텍스트로 시각적 설명자를 강화할 수 있다. 신뢰도 분류기(c)는 추론을 중지할지 여부를 결정할 수 있다. 확신할 수 있는 포인트가 거의 없으면 추론은 다음 레이어로 진행되지만 확실하게 일치하지 않는 포인트는 제거한다. LightGlue는 확신 상태에 도달하는 경우, 한 쌍의 유사성(similarity) 및 단일의 일치성(matchability)을 기반으로 포인트 간의 할당(assignment)을 예측한다.
- [0056] 일 실시예로, 인공지능 모델은, 추출한 특징점들로부터 2차원 이미지(2D Image)와 기저장된 3차원 비교 이미지(3D Image)를 매칭하는 매칭 포인트들을 포함하는 3차원 변환 정보를 추정하도록 학습된 것일 수 있다.
- [0057] 일 실시예로, 인공지능 모델은, 인코더(Keypoint Encoder), 그래프 신경망(Multiplex Graph Neural Network) 및 매칭 레이어(Optimal Matching Layer)를 포함할 수 있다.
- [0058] 일 실시예로, 인코더(Keypoint Encoder)는 특징점의 위치와 외관 정보를 인코딩하는 역할을 수행하며, 인코딩된 정보는 그래프 신경망에 의해 처리되어 매칭 과정에서 중요한 역할을 할 수 있다.
- [0059] 일 실시예로, 그래프 신경망(Multiplex Graph Neural Network)은 두 이미지 간의 특징점들 사이의 관계를 모델링하는 고급 네트워크일 수 있다. 그래프 신경망은 어떤 특징점들이 서로 매칭될 가능성이 높은지를 효과적으로 판단할 수 있다.
- [0060] 일 실시예로, 매칭 레이어(Optimal Matching Layer)는 그래프 신경망에서 생성된 정보를 기반으로 최적의 매칭 쌍을 결정하는 결정적인 레이어일 수 있다. 이 과정에서 매칭이 불가능한 특징점들은 자동으로 제외될 수 있다.
- [0061] 일 실시예에 따른 포즈 추정 장치(100)는, 동작 480에서, 2차원 이미지(2D Image)에 대응하는 3차원 변환 정보(예: PnP output, 2D camera matrix)를 추정할 수 있다.

[0062] 일 실시예로, 포즈 추정 장치(100)는 2D 카메라(210)와 기저장된 3차원 비교 이미지(3D Image)에 대응하는 3D 카메라(미도시) 사이의 상관 관계를 3차원 변환 정보로 추정할 수 있다.

[0063] 일 실시예로, 포즈 추정 장치(100)는 기저장된 3차원 비교 이미지(3D Image)에 대응하는 3D 카메라(미도시)와 엔드 이펙터(239) 사이의 상대적 위치에 대응하는 제1 변환 매트릭스 및 2차원 이미지(2D Image)에 대응하는 2D 카메라(210)와 로봇의 베이스(235) 사이의 상대적 위치에 대응하는 제2 변환 매트릭스에 기초하여, 3차원 변환 정보를 추정할 수 있다.

[0064] 일 실시예로, 포즈 추정 장치(100)는 변환 매트릭스에 기초하여 로봇의 좌표계와 카메라의 좌표계를 캘리브레이션할 수 있다. 예를 들어, 엔드 이펙터(239)와 카메라 사이의 상대적 위치 및 방향을 구하기 위한 변환 매트릭스(T)는 하기와 같은 수학적식일 수 있다.

**수학식 1**

[0065] 
$${}^{endeffector}T_{camera} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

[0066] 상기의 변환 매트릭스( ${}^{endeffector}T_{camera}$ )는 동차 좌표계를 사용하는 4X4 행렬일 수 있다. 상위 3X3 행렬은 각각의 축 방향에 따른 회전(rotation)을 나타내는 rotation vector일 수 있다. 마지막 열의 상위 3개 요소는 이동(translation)을 나타내는 상대적 위치일 수 있다.

[0067] 일 실시예로, 베이스(235)와 엔드 이펙터(239) 사이의 변환 매트릭스( ${}^{base}T_{endeffector}$ )는 엔드 이펙터(239)에 툴을 설치한 로봇 장치(230)로부터 획득될 수 있다.

[0068] 일 실시예로, 포즈 추정 장치(100)는 상기 수학식 1에 대응하는 변환 매트릭스( ${}^{endeffector}T_{camera}$ ) 및 베이스(235)와 엔드 이펙터(239) 사이의 변환 매트릭스( ${}^{base}T_{endeffector}$ )에 기초하여, 기저장된 3차원 비교 이미지(3D Image)에 대응하는 3D 카메라(미도시)와 엔드 이펙터(239) 사이의 상대적 위치에 대응하는 제1 변환 매트릭스( ${}^{3Dcamera}T_{endeffector}$ )를 하기의 수학적식으로 추정할 수 있다.

**수학식 2**

[0069] 
$${}^{3Dcamera}T_{endeffector} = ({}^{base}T_{3Dcamera})^{-1} \times {}^{base}T_{endeffector} = {}^{3Dcamera}T_{base} \times {}^{base}T_{endeffector}$$

[0070] 일 실시예로, 포즈 추정 장치(100)는 상기 수학식 1에 대응하는 변환 매트릭스( ${}^{endeffector}T_{camera}$ ) 및 베이스(235)와 엔드 이펙터(239) 사이의 변환 매트릭스( ${}^{base}T_{endeffector}$ )에 기초하여, 2차원 이미지(2D Image)에 대응하는 2D 카메라(210)와 로봇의 베이스(235) 사이의 상대적 위치에 대응하는 제2 변환 매트릭스( ${}^{base}T_{2Dcamera}$ )를 추정할 수 있다.

[0071] 일 실시예로, 포즈 추정 장치(100)는 매칭 포인트들에 기초하여, PnP(Perspective N Points) 방식을 사용함으로써 기저장된 3차원 비교 이미지(3D Image)에 대응하는 3D 카메라(미도시)에 대비한 2차원 이미지(2D Image)에 대응하는 2D 카메라(210)의 자세 데이터를 추정할 수 있다.

[0072] 일 실시예로, PnP 방식은 2차원 이미지(2D Image)의 특징점들(예: matching keypoint)과 기저장된 3차원 비교 이미지(3D Image)의 매칭 포인트(예: 3D camera PointCloud)를 서로 연결함으로써, 2D 카메라(210)의 자세 데

이터를 추정할 수 있다. 예를 들어, 2D 카메라(210)의 자세 데이터에는 2D 카메라(210)의 위치 및/또는 방향이 포함될 수 있다.

- [0073] 일 실시예로, 포즈 추정 장치(100)는 2D 카메라(210)의 내부 매트릭스에 기초하여, 2차원 이미지(2D Image)의 특징점들을 3차원 공간으로 이동할 수 있고, 3차원 공간에서의 특징점들에 기초하여 2D 카메라(210)의 자세 데이터를 추정할 수 있다.
- [0074] 일 실시예로, 포즈 추정 장치(100)는 적은 개수의 포인트들로 정확한 포즈를 추정할 수 있는 EPnP(Efficient PnP) 방식을 이용할 수 있다.
- [0075] 일 실시예에 따른 포즈 추정 장치(100)는, RANSAC(random sample consensus) 기법을 이용함으로써, 매칭 포인트들에 기초하여 적어도 하나 이상의 모델을 추정하고, 추정한 모델 중 최적 모델을 선택하도록 학습된 것인,
- [0076] 일 실시예로, RANSAC 기법은 데이터 중 이상값이 있어도 견고한 모델을 추정할 수 있는 반복적인 방법일 수 있다.
- [0077] 예를 들어, RANSAC 기법에 따르면, 포즈 추정 장치(100)는 데이터 세트(예: 특징점-매칭 포인트)에서 무작위로 일부를 선별하고(샘플링), 선별된 일부에 기초하여 모델(예: 2D 카메라(210)의 포즈)을 추정하고, 추정한 모델을 이용하여 전체 데이터를 대상으로 각 포인트들에 대한 오차를 계산하고, 미리 정의된 임계값보다 작은 오차를 갖는 포인트를 내부점(inliers)으로 판단하고(적합도 평가), 이러한 과정을 반복하여 가장 많은 내부점(inliers)을 얻을 수 있는 최적의 모델을 추정할 수 있다(최적 모델 선택). 또한, 최종적으로 추정된 최적 모델에 기초하여, 내부점들만을 사용하여, 모델을 다시 조정 및 정제할 수 있다.
- [0078] 일 실시예에 따른 포즈 추정 장치(100)는, 동작 490에서, 추정한 3차원 변환 정보에 기초하여, 엔드 이펙터(239)의 포즈를 추정할 수 있다.
- [0079] 일 실시예로, 포즈 추정 장치(100)는 추정한 3차원 변환 정보에 기초하여, 2D 카메라(210)로부터 획득한 2차원 이미지(2D Image)에 대응하는 3차원 데이터를 획득할 수 있다. 일 실시예로, 포즈 추정 장치(100)는 획득한 3차원 데이터에 기초하여 엔드 이펙터(239)의 포즈를 추정할 수 있다.
- [0080] 따라서, 일 실시예에 따른 포즈 추정 장치(100)에 따르면, 고해상도의 이미지를 제공 가능한 2D 카메라(210)의 장점과 깊이 정보를 제공 가능한 3D 카메라(미도시)의 장점을 결합함으로써, 고해상도 이미지와 깊이 정보를 함께 획득할 수 있다.
- [0081] 또한, 일 실시예에 따른 포즈 추정 장치(100)에 따르면, 2차원 이미지(2D Image)와 3차원 데이터를 함께 사용함으로써, 물체 인식과 추적의 정확도를 크게 향상시킬 수 있다.
- [0082] 또한, 2D 카메라(210)는 조명 변화에 민감하나, 3D 카메라(미도시)는 조명의 영향을 상대적으로 적게 받는 점에서, 일 실시예에 따른 포즈 추정 장치(100)에 따르면, 다양한 조명 조건 하에서도 안정적인 성능을 확보할 수 있다.
- [0083] 또한, 일 실시예에 따른 포즈 추정 장치(100)에 따르면, 필요에 따라 3차원 데이터를 보완적으로 활용함으로써 전체 시스템의 비용 효율성을 높일 수 있다.
- [0084] 상기 설명된 포즈 추정 장치(100)에서 포즈 추정 방법은 컴퓨터에 의해 실행되는 컴퓨터 판독가능 기록매체에 저장된 컴퓨터 프로그램 또는 컴퓨터에 의해 실행 가능한 명령어들을 포함하는 기록매체의 형태로도 구현될 수 있다. 또한, 상기 설명된 포즈 추정 장치(100)에서 포즈 추정 방법은 컴퓨터에 의해 실행되는 컴퓨터 판독가능 기록매체에 저장된 컴퓨터 프로그램의 형태로도 구현될 수 있다.
- [0085] 컴퓨터 판독 가능 기록매체는 컴퓨터에 의해 액세스될 수 있는 임의의 가용 매체일 수 있고, 휘발성 및 비휘발성 매체, 분리형 및 비분리형 매체를 모두 포함한다. 또한, 컴퓨터 판독가능 기록매체는 컴퓨터 저장 매체를 포함할 수 있다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 모두 포함한다.
- [0086] 본원 명세서 내에 기재된 구성요소에 의해 실현되는 기능은 해당 기재된 기능을 실현하도록 프로그램된 범용 프로세서, 특정 용도 프로세서, 집적회로, ASICs(Application Specific Integrated Circuits), CPU(Central Processing Unit), 회로 및/또는 이들 조합을 포함하는 프로세싱 회로(processing circuitry)에서 구현되어도 된다. 프로세서는 트랜지스터나 기타 회로를 포함하며, 회로 또는 프로세싱 회로로 간주된다. 프로세서는 메모

리에 저장된 프로그램을 실행하는 프로그램된 프로세서(programmed processor)여도 좋다.

[0087] 본원 명세서에서, 회로, 부, 유닛, 수단은 기재된 기능을 실현하도록 프로그램된 하드웨어 또는 실행하는 하드웨어이다. 해당 하드웨어는 본원 명세서에 개시된 모든 하드웨어 또는 해당 기재된 기능을 실현하도록 프로그램되거나 실행하는 것으로 알려진 임의의 하드웨어라도 무방하다.

[0088] 해당 하드웨어가 회로 타입이라고 간주되는 프로세서인 경우, 해당 회로, 해당 부, 수단 또는 유닛은 하드웨어와 해당 하드웨어 및 또는 프로세서를 구성하기 위해 이용되는 소프트웨어의 조합이다.

[0089] 전술한 본 개시의 설명은 예시를 위한 것이며, 본 개시가 속하는 기술분야의 통상의 지식을 가진 자는 본 개시의 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다. 예를 들어, 단일형으로 설명되어 있는 각 구성 요소는 분산되어 실시될 수도 있으며, 마찬가지로 분산된 것으로 설명되어 있는 구성 요소들도 결합된 형태로 실시될 수 있다.

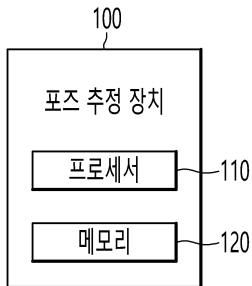
[0090] 본 개시의 범위는 상기 상세한 설명보다는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 개시의 범위에 포함되는 것으로 해석되어야 한다.

**부호의 설명**

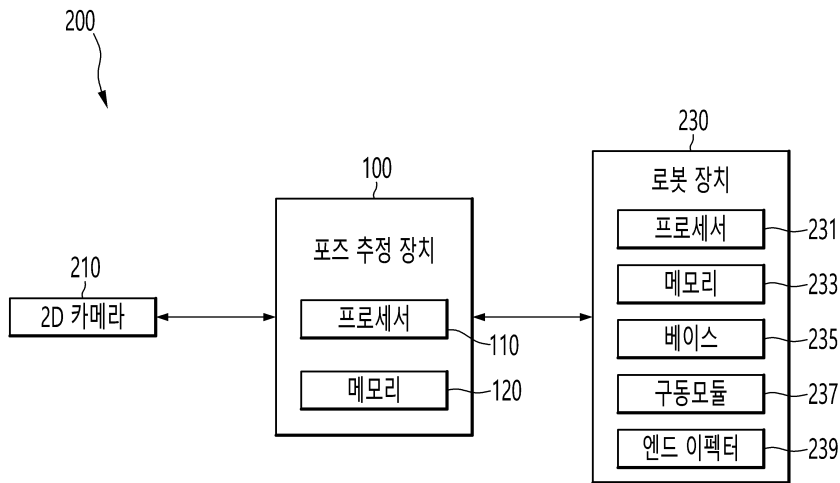
- [0091] 100 : 포즈 추정 장치
- 200 : 로봇 제어 시스템
- 210 : 2D 카메라
- 230 : 로봇 장치

**도면**

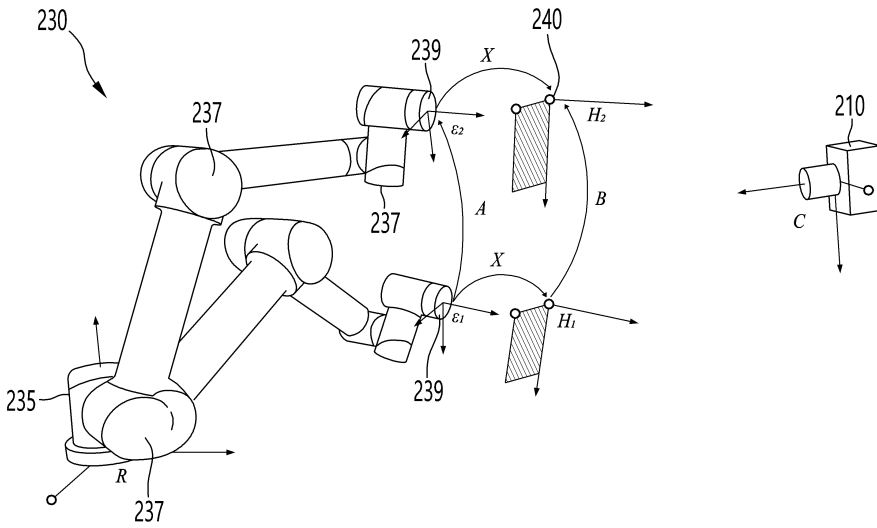
**도면1**



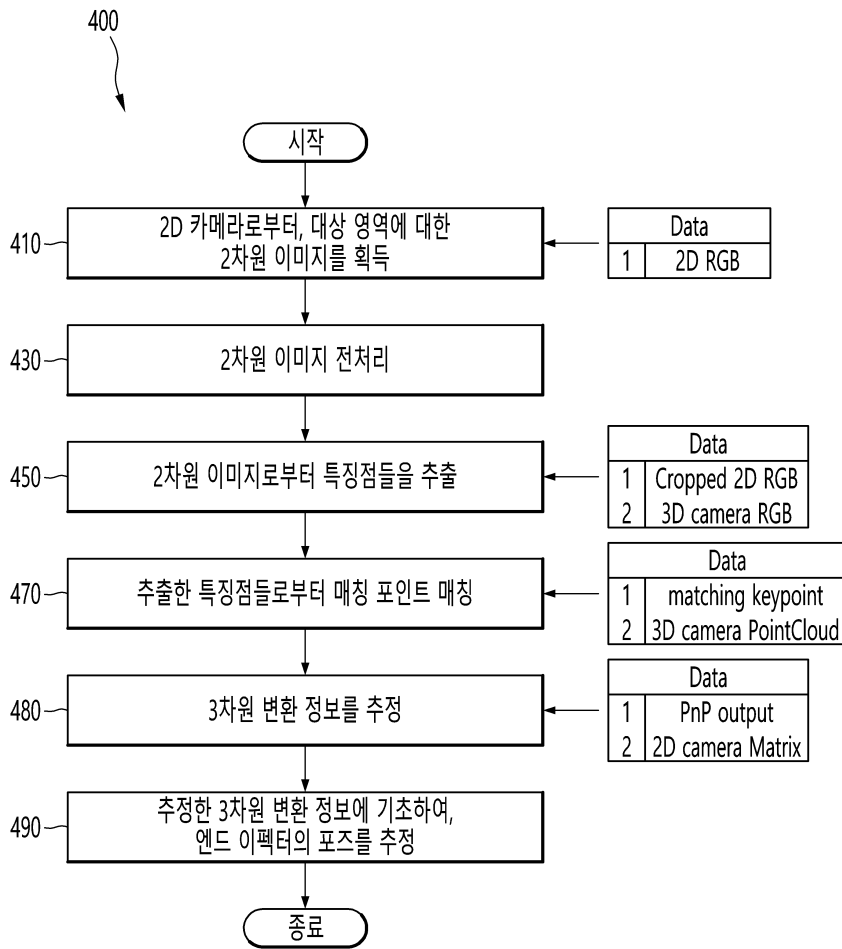
도면2



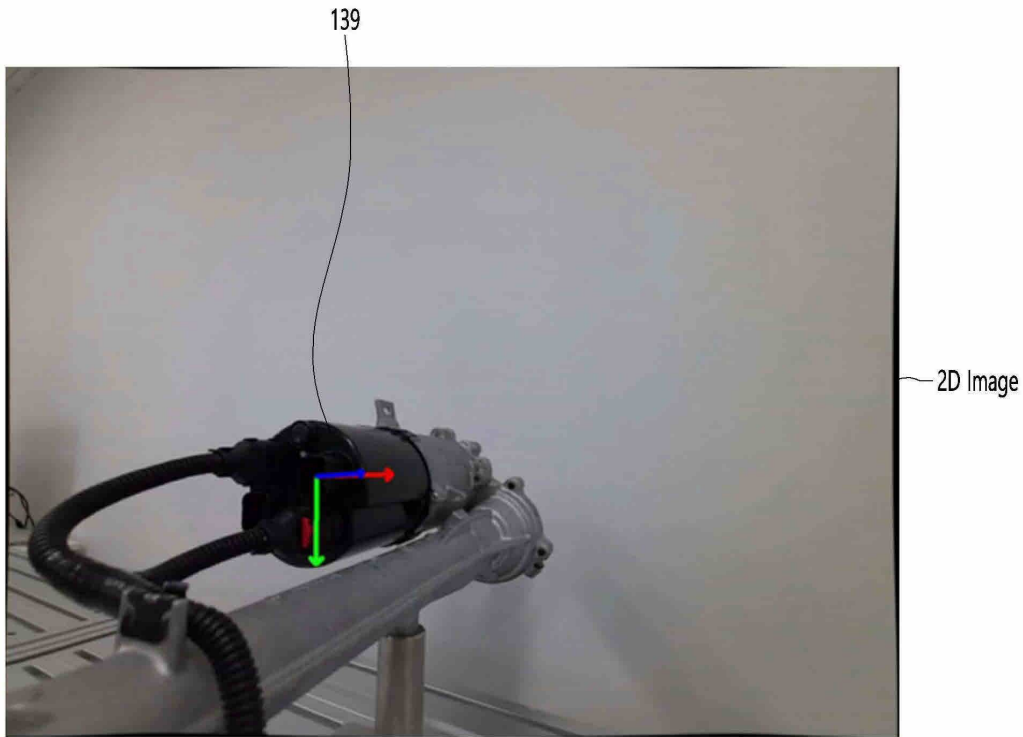
도면3



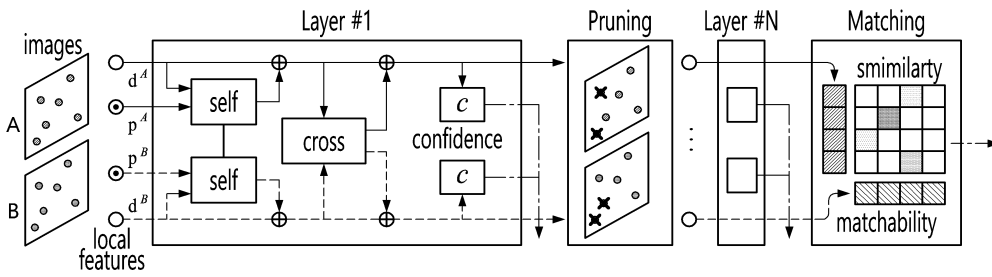
도면4



도면5



도면6



도면7

