



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0089466
(43) 공개일자 2023년06월20일

(51) 국제특허분류(Int. Cl.)

G06V 40/20 (2022.01) G06N 3/04 (2023.01)
G06T 3/40 (2006.01) G06T 7/90 (2017.01)
G06V 10/12 (2022.01) G06V 10/764 (2022.01)
G06V 10/80 (2022.01)

(52) CPC특허분류

G06V 40/20 (2022.01)
G06N 3/045 (2023.01)

(21) 출원번호 10-2021-0178108

(22) 출원일자 2021년12월13일

심사청구일자 2021년12월13일

(71) 출원인

조선대학교산학협력단

광주광역시 동구 필문대로 309 (서석동)

(72) 발명자

박근창

광주광역시 서구 풍암2로 66 금호타운 202동 703호

변영현

전라남도 장성군 장성읍 오동촌길 16 부강아트빌 303호

박지원

광주광역시 북구 용봉로126번길 37-2

(74) 대리인

특허법인지원

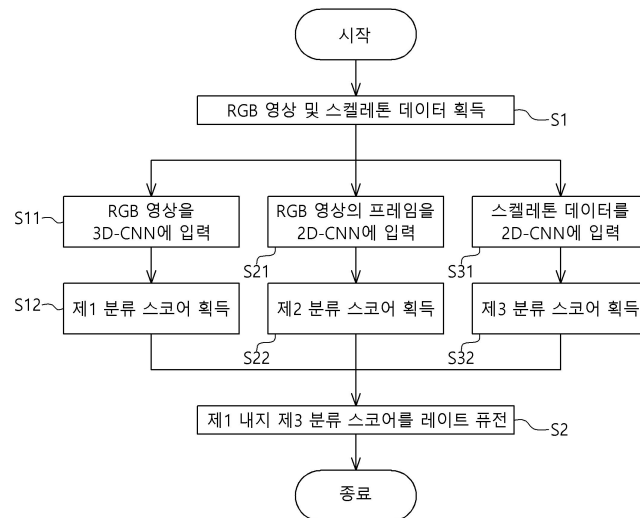
전체 청구항 수 : 총 12 항

(54) 발명의 명칭 양상불 기반의 신경망을 이용한 행동 인식 방법

(57) 요약

본 발명의 일 실시예에 따른 양상불 기반의 신경망을 이용한 행동 인식 방법은 RGB 영상과 상기 RGB 영상 내 인체의 스켈레톤 데이터를 획득하는 단계, 상기 RGB 영상을 3차원 컨볼루션 신경망에 입력하여 제1 분류 스코어를 획득하는 단계, 상기 RGB 영상의 각 프레임 및 상기 스켈레톤 데이터를 서로 다른 2차원 컨볼루션 신경망에 입력하여 제2 및 제3 분류 스코어를 획득하는 단계 및 상기 제1 내지 제3 분류 스코어를 레이트 퓨전하는 단계를 포함하는 것을 특징으로 한다.

대표도 - 도1



(52) CPC특허분류

G06T 3/4046 (2013.01)

G06T 7/90 (2017.01)

G06V 10/12 (2023.01)

G06V 10/764 (2023.01)

G06V 10/811 (2023.01)

G06T 2207/20044 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711138632
과제번호	S1601-20-1041
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신산업진흥원
연구사업명	인공지능 중심 산업 융합 집적단지 조성사업
연구과제명	AI 기반 다중센싱을 이용한 Brain-Body 인터페이스 기술 개발
기 여 율	1/2
과제수행기관명	조선대학교 산학협력단
연구기간	2020.04.01 ~ 2024.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	2021812
과제번호	2021-0812
부처명	과학기술정보통신부
과제관리(전문)기관명	인공지능산업융합사업단
연구사업명	인공지능중심산업융합집적단지조성사업 AI융합대학지원사업
연구과제명	AI융합대학지원사업
기 여 율	1/2
과제수행기관명	조선대학교 산학협력단
연구기간	2021.07.01 ~ 2021.12.31

명세서

청구범위

청구항 1

RGB 영상과 상기 RGB 영상 내 인체의 스켈레톤 데이터를 획득하는 단계;

상기 RGB 영상을 3차원 컨볼루션 신경망에 입력하여 제1 분류 스코어를 획득하는 단계;

상기 RGB 영상의 각 프레임 및 상기 스켈레톤 데이터를 서로 다른 2차원 컨볼루션 신경망에 입력하여 제2 및 제3 분류 스코어를 획득하는 단계; 및

상기 제1 내지 제3 분류 스코어를 레이트 퓨전하는 단계를 포함하는

양상블 기반의 신경망을 이용한 행동 인식 방법.

청구항 2

제1항에 있어서,

상기 RGB 영상과 상기 스켈레톤 데이터를 획득하는 단계는

키넥트(Kinect) 센서로부터 상기 RGB 영상과 상기 스켈레톤 데이터를 수신하는 단계를 포함하는

양상블 기반의 신경망을 이용한 행동 인식 방법.

청구항 3

제1항에 있어서,

상기 RGB 영상과 상기 스켈레톤 데이터를 획득하는 단계는

칼라 센서로부터 수신된 RGB 영상에서 상기 스켈레톤 데이터를 추출하는 단계를 포함하는

양상블 기반의 신경망을 이용한 행동 인식 방법.

청구항 4

제1항에 있어서,

상기 스켈레톤 데이터는 상기 RGB 영상의 연속되는 프레임에 대응하는 스켈레톤 시퀀스를 포함하는

양상블 기반의 신경망을 이용한 행동 인식 방법.

청구항 5

제1항에 있어서,

상기 RGB 영상을 상기 3차원 컨볼루션 신경망에 입력하는 단계는

상기 RGB 영상을 구성하는 연속된 프레임 묶음을 상기 3차원 컨볼루션 신경망에 입력하는 단계를 포함하는

양상블 기반의 신경망을 이용한 행동 인식 방법.

청구항 6

제1항에 있어서,

상기 제2 분류 스코어를 획득하는 단계는

상기 RGB 영상의 각 프레임을 2차원 컨볼루션 신경망에 입력하는 단계와,

상기 2차원 컨볼루션 신경망을 통해 추출된 특징의 시퀀스를 순환 신경망에 입력하여 상기 제2 분류 스코어를 획득하는 단계를 포함하는

양상블 기반의 신경망을 이용한 행동 인식 방법.

청구항 7

제6항에 있어서,

상기 순환 신경망은 LSTM(Long Short-Term Memory) 모델인

양상블 기반의 신경망을 이용한 행동 인식 방법.

청구항 8

제1항에 있어서,

상기 제3 분류 스코어를 획득하는 단계는

상기 스켈레톤 데이터를 RGB 이미지로 변환하는 단계와,

상기 RGB 이미지를 상기 2차원 컨볼루션 신경망에 입력하여 상기 제3 분류 스코어를 획득하는 단계를 포함하는

양상블 기반의 신경망을 이용한 행동 인식 방법.

청구항 9

제8항에 있어서,

상기 스켈레톤 데이터를 상기 RGB 이미지로 변환하는 단계는

상기 스켈레톤 데이터의 3차원 좌표 값을 R, G, B 색상에 매핑하여 상기 RGB 이미지를 생성하는 단계를 포함하는

양상블 기반의 신경망을 이용한 행동 인식 방법.

청구항 10

제8항에 있어서,

상기 스켈레톤 데이터를 상기 RGB 이미지로 변환하는 단계는

상기 스켈레톤 데이터를, 상기 스켈레톤 데이터의 수와 상기 RGB 영상의 프레임 수의 두 좌표축을 갖는 상기 RGB 이미지로 변환하는 단계를 포함하는

양상블 기반의 신경망을 이용한 행동 인식 방법.

청구항 11

제1항에 있어서,

상기 제1 내지 제3 분류 스코어를 레이트 퓨전하는 단계는

상기 제1 내지 제3 분류 스코어의 평균하여 최종 스코어 행렬을 산출하는 단계를 포함하는

양상블 기반의 신경망을 이용한 행동 인식 방법.

청구항 12

제1항에 있어서,

상기 제1 내지 제3 분류 스코어를 레이트 퓨전하는 단계는

상기 제1 내지 제3 분류 스코어를 요소별로(elementwise) 곱하여 최종 스코어 행렬을 산출하는 단계를 포함하는

양상블 기반의 신경망을 이용한 행동 인식 방법.

발명의 설명

기술 분야

[0001] 본 발명은 RGB 영상과 스켈레톤 데이터에 3-스트림 양상블 모델을 적용하여 행동 인식을 수행하는 방법에 관한 것이다.

배경 기술

[0003] 인간의 행위 의도를 파악하고 그에 맞는 서비스를 제공하기 위하여, 컴퓨터 비전 또는 인공지능을 이용한 행동 인식에 관한 연구가 활발히 이루어지고 있다.

[0004] 컴퓨터 비전 분야에서 행동 인식에 관한 연구는 수십 년간 이루어졌으나 객체의 오버랩(overlap), 객체의 외형 변화, 비 강체(non-rigid body)의 움직임, 정적 데이터와 동적 데이터의 혼재 등으로 인해 행동 인식의 신뢰성을 확보하는데 어려움이 있었다.

[0005] 이러한 한계를 극복하기 위하여 최근 행동 인식에 딥러닝 신경망(Deep learning neural network), 대표적으로는 컨볼루션 신경망(Convolutional Neural Network; CNN)이 적용되고 있다.

[0006] 그러나, CNN의 경우 특정 프레임의 이미지에 대해서만 행동 분석을 수행하므로, 과거의 정보에 기초한 연속적인 행동 인식에 활용할 수 없다는 한계가 있다. 또한, 단일 신경망을 이용한 행동 인식의 신뢰성을 확보하기 위해서는 방대한 학습 데이터가 필요하다는 문제가 있다.

발명의 내용

해결하려는 과제

[0008] 본 발명은 RGB 영상과 스켈레톤 데이터에 3-스트림 양상블 모델을 적용하여 행동 인식을 수행하는 것을 목적으로 한다.

[0009] 본 발명의 목적들은 이상에서 언급한 목적으로 제한되지 않으며, 언급되지 않은 본 발명의 다른 목적 및 장점들은 하기의 설명에 의해서 이해될 수 있고, 본 발명의 실시예에 의해 보다 분명하게 이해될 것이다. 또한, 본 발명의 목적 및 장점들은 특허 청구 범위에 나타낸 수단 및 그 조합에 의해 실현될 수 있음을 쉽게 알 수 있을 것이다.

과제의 해결 수단

- [0011] 진술한 목적을 달성하기 위한 본 발명의 일 실시예에 따른 양상블 기반의 신경망을 이용한 행동 인식 방법은 RGB 영상과 상기 RGB 영상 내 인체의 스켈레톤 데이터를 획득하는 단계, 상기 RGB 영상을 3차원 컨볼루션 신경망에 입력하여 제1 분류 스코어를 획득하는 단계, 상기 RGB 영상의 각 프레임 및 상기 스켈레톤 데이터를 서로 다른 2차원 컨볼루션 신경망에 입력하여 제2 및 제3 분류 스코어를 획득하는 단계 및 상기 제1 내지 제3 분류 스코어를 레이트 퓨전하는 단계를 포함하는 것을 특징으로 한다.
- [0012] 일 실시예에서, 상기 RGB 영상과 상기 스켈레톤 데이터를 획득하는 단계는 키넥트(Kinect) 센서로부터 상기 RGB 영상과 상기 스켈레톤 데이터를 수신하는 단계를 포함하는 것을 특징으로 한다.
- [0013] 일 실시예에서, 상기 RGB 영상과 상기 스켈레톤 데이터를 획득하는 단계는 칼라 센서로부터 수신된 RGB 영상에서 상기 스켈레톤 데이터를 추출하는 단계를 포함하는 것을 특징으로 한다.
- [0014] 일 실시예에서, 상기 스켈레톤 데이터는 상기 RGB 영상의 연속되는 프레임에 대응하는 스켈레톤 시퀀스를 포함하는 것을 특징으로 한다.
- [0015] 일 실시예에서, 상기 RGB 영상을 상기 3차원 컨볼루션 신경망에 입력하는 단계는 상기 RGB 영상을 구성하는 연속된 프레임 묶음을 상기 3차원 컨볼루션 신경망에 입력하는 단계를 포함하는 것을 특징으로 한다.
- [0016] 일 실시예에서, 상기 제2 분류 스코어를 획득하는 단계는 상기 RGB 영상의 각 프레임을 2차원 컨볼루션 신경망에 입력하는 단계와, 상기 2차원 컨볼루션 신경망을 통해 추출된 특징의 시퀀스를 순환 신경망에 입력하여 상기 제2 분류 스코어를 획득하는 단계를 포함하는 것을 특징으로 한다.
- [0017] 일 실시예에서, 상기 순환 신경망은 LSTM(Long Short-Term Memory) 모델인 것을 특징으로 한다.
- [0018] 일 실시예에서, 상기 제3 분류 스코어를 획득하는 단계는 상기 스켈레톤 데이터를 RGB 이미지로 변환하는 단계와, 상기 RGB 이미지를 상기 2차원 컨볼루션 신경망에 입력하여 상기 제3 분류 스코어를 획득하는 단계를 포함하는 것을 특징으로 한다.
- [0019] 일 실시예에서, 상기 스켈레톤 데이터를 상기 RGB 이미지로 변환하는 단계는 상기 스켈레톤 데이터의 3차원 좌표 값을 R, G, B 색상에 매핑하여 상기 RGB 이미지를 생성하는 단계를 포함하는 것을 특징으로 한다.
- [0020] 일 실시예에서, 상기 스켈레톤 데이터를 상기 RGB 이미지로 변환하는 단계는 상기 스켈레톤 데이터, 상기 스켈레톤 데이터의 수와 상기 RGB 영상의 프레임 수의 두 좌표축을 갖는 상기 RGB 이미지로 변환하는 단계를 포함하는 것을 특징으로 한다.
- [0021] 일 실시예에서, 상기 제1 내지 제3 분류 스코어를 레이트 퓨전하는 단계는 상기 제1 내지 제3 분류 스코어의 평균하여 최종 스코어 행렬을 산출하는 단계를 포함하는 것을 특징으로 한다.
- [0022] 일 실시예에서, 상기 제1 내지 제3 분류 스코어를 레이트 퓨전하는 단계는 상기 제1 내지 제3 분류 스코어를 요소별로(elementwise) 곱하여 최종 스코어 행렬을 산출하는 단계를 포함하는 것을 특징으로 한다.

발명의 효과

- [0024] 본 발명은 3-스트림 양상블 모델을 통해 서로 다른 특징을 갖는 RGB 영상과 스켈레톤 데이터로부터 최적의 정보를 효율적으로 추출 및 병합함으로써 신경망을 이용한 행동 인식의 신뢰성을 향상시킬 수 있다.
- [0025] 상술한 효과와 더불어 본 발명의 구체적인 효과는 이하 발명을 실시하기 위한 구체적인 사항을 설명하면서 함께 기술한다.

도면의 간단한 설명

- [0027] 도 1은 본 발명의 일 실시예에 따른 양상블 기반의 신경망을 이용한 행동 인식 방법을 도시한 순서도.
- 도 2는 RGB 영상과 해당 영상 내 인체의 스켈레톤 데이터를 예시적으로 도시한 도면.
- 도 3은 스켈레톤 데이터를 설명하기 위한 도면.
- 도 4는 RGB 영상의 각 프레임별 스켈레톤 시퀀스를 도시한 도면.

도 5는 3차원 컨볼루션 신경망에 RGB 영상이 입력되는 모습을 도시한 도면.

도 6은 2차원 컨볼루션 신경망과 순환 신경망이 연결된 구조를 도시한 도면.

도 7은 2차원 컨볼루션 신경망과 LSTM이 연결된 구조에 RGB 영상의 각 프레임이 입력되는 모습을 도시한 도면.

도 8은 스켈레톤 데이터를 RGB 이미지로 변환하고 RGB 이미지를 2차원 컨볼루션 신경망에 입력하는 모습을 도시한 도면.

도 9는 스켈레톤 데이터를 RGB 이미지로 변환하는 과정을 도시한 도면.

도 10은 스켈레톤 데이터의 3차원 좌표 값을 R, G, B 색상에 매핑함으로써 RGB 이미지를 생성하는 과정을 도시한 도면.

발명을 실시하기 위한 구체적인 내용

- [0028] 진술한 목적, 특징 및 장점은 첨부된 도면을 참조하여 상세하게 후술되며, 이에 따라 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 본 발명의 기술적 사상을 용이하게 실시할 수 있을 것이다. 본 발명을 설명함에 있어서 본 발명과 관련된 공지 기술에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 상세한 설명을 생략한다. 이하, 첨부된 도면을 참조하여 본 발명에 따른 바람직한 실시예를 상세히 설명하기로 한다. 도면에서 동일한 참조부호는 동일 또는 유사한 구성요소를 가리키는 것으로 사용된다.
- [0029] 본 명세서에서 제1, 제2 등이 다양한 구성요소들을 서술하기 위해서 사용되나, 이들 구성요소들은 이들 용어에 의해 제한되지 않음은 물론이다. 이들 용어들은 단지 하나의 구성요소를 다른 구성요소와 구별하기 위하여 사용하는 것으로, 특별히 반대되는 기재가 없는 한, 제1 구성요소는 제2 구성요소일 수도 있음은 물론이다.
- [0030] 또한, 본 명세서에서 "상부 (또는 하부)" 또는 구성요소의 "상 (또는 하)"에 임의의 구성이 배치된다는 것은, 임의의 구성이 상기 구성요소의 상면 (또는 하면)에 접하여 배치되는 것뿐만 아니라, 상기 구성요소와 상기 구성요소 상에 (또는 하에) 배치된 임의의 구성 사이에 다른 구성이 개재될 수 있음을 의미할 수 있다.
- [0031] 또한, 본 명세서에서 어떤 구성요소가 다른 구성요소에 "연결", "결합" 또는 "접속"된다고 기재된 경우, 상기 구성요소들은 서로 직접적으로 연결되거나 또는 접속될 수 있지만, 각 구성요소 사이에 다른 구성요소가 "개재"되거나, 각 구성요소가 다른 구성요소를 통해 "연결", "결합" 또는 "접속"될 수도 있는 것으로 이해되어야 할 것이다.
- [0032] 또한, 본 명세서에서 사용되는 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 출원에서, "구성된다" 또는 "포함한다" 등의 용어는 명세서 상에 기재된 여러 구성 요소들, 또는 여러 단계들을 반드시 모두 포함하는 것으로 해석되지 않아야 하며, 그 중 일부 구성 요소들 또는 일부 단계들은 포함되지 않을 수도 있고, 또는 추가적인 구성 요소 또는 단계들을 더 포함할 수 있는 것으로 해석되어야 한다.
- [0033] 또한, 본 명세서에서, "A 및/또는 B" 라고 할 때, 이는 특별한 반대되는 기재가 없는 한, A, B 또는 A 및 B 를 의미하며, "C 내지 D" 라고 할 때, 이는 특별한 반대되는 기재가 없는 한, C 이상이고 D 이하인 것을 의미한다.
- [0034] 본 발명은 RGB 영상과 스켈레톤 데이터에 3-스트림 앙상블 모델을 적용하여 행동 인식을 수행하는 방법에 관한 것이다. 구체적으로, 본 발명은 RGB 영상 및 스켈레톤 데이터를 세 개의 신경망 네트워크 스트림에 입력하고, 그 출력을 앙상블하여 행동 인식을 수행하는 방법에 관한 것이다. 이하, 도 1 내지 도 10을 참조하여, 본 발명의 일 실시예에 따른 앙상블 기반의 신경망을 이용한 행동 인식 방법을 구체적으로 설명하도록 한다.
- [0035] 도 1은 본 발명의 일 실시예에 따른 앙상블 기반의 신경망을 이용한 행동 인식 방법을 도시한 순서도이다.
- [0036] 도 2는 RGB 영상과 해당 영상 내 인체의 스켈레톤 데이터를 예시적으로 도시한 도면이고, 도 3은 스켈레톤 데이터를 설명하기 위한 도면이며, 도 4는 RGB 영상의 각 프레임별 스켈레톤 시퀀스를 도시한 도면이다.
- [0037] 도 5는 3차원 컨볼루션 신경망에 RGB 영상이 입력되는 모습을 도시한 도면이다.
- [0038] 도 6은 2차원 컨볼루션 신경망과 순환 신경망이 연결된 구조를 도시한 도면이고, 도 7은 2차원 컨볼루션 신경망과 LSTM이 연결된 구조에 RGB 영상의 각 프레임이 입력되는 모습을 도시한 도면이다.
- [0039] 도 8은 스켈레톤 데이터를 RGB 이미지로 변환하고 RGB 이미지를 2차원 컨볼루션 신경망에 입력하는 모습을 도시

한 도면이다.

- [0040] 도 9는 스켈레톤 데이터를 RGB 이미지로 변환하는 과정을 도시한 도면이고, 도 10은 스켈레톤 데이터의 3차원 좌표 값을 R, G, B 색상에 매핑함으로써 RGB 이미지를 생성하는 과정을 도시한 도면이다.
- [0041] 도 1을 참조하면, 본 발명의 일 실시예에 따른 양상불 기반의 신경망을 이용한 행동 인식 방법(이하, 행동 인식 방법)은 RGB 영상(100) 및 스켈레톤 데이터(200)를 획득하는 단계(S1), RGB 영상(100)을 3D-CNN(3 Dimensional Convolutional neural network)에 입력하여(S11) 제1 분류 스코어를 획득하는 단계(S12), RGB 영상(100)의 프레임(110)을 2D-CNN에 입력하여(S21) 제2 분류 스코어를 획득하는 단계(S22), 스켈레톤 데이터(200)를 2D-CNN에 입력하여(S31) 제3 분류 스코어를 획득하는 단계(S32)와, 제1 내지 제3 분류 스코어를 레이트 퓨전(late-fusion)하는 단계(S2)를 포함할 수 있다.
- [0042] 다만, 도 1에 도시된 행동 인식 방법은 일 실시예에 따른 것이고, 발명을 이루는 각 단계들이 도 1에 도시된 실시예에 한정되는 것은 아니며, 필요에 따라 일부 단계가 부가, 변경 또는 삭제될 수 있다.
- [0043] 한편, 도 1에 도시된 행동 인식 방법은 프로세서에 의해 수행될 수 있으며, 프로세서는 발명의 동작을 수행하기 위하여 CPU(central processing unit), GPU(graphic processing unit), ASICs(application specific integrated circuits), DSPs(digital signal processors), DSPDs(digital signal processing devices), PLDs(programmable logic devices), FPGAs(field programmable gate arrays), 제어기(controller), 마이크로 컨트롤러(micro-controllers) 중 적어도 하나의 물리적인 요소를 포함할 수 있다.
- [0044] 프로세서는 RGB 영상(100)과 RGB 영상(100) 내 인체의 스켈레톤 데이터(200)를 획득할 수 있다(S1).
- [0045] 도 2를 참조하면, RGB 영상(100)은 픽셀별로 Red, Blue, Green의 색상값을 갖는 영상으로, 다수의 프레임(110)으로 구성될 수 있다. 다시 말해, RGB 영상(100)은 각 프레임(110)에 대응하는 이미지를 포함할 수 있다.
- [0046] 한편, 도 2 및 도 3을 함께 참조하면 스켈레톤 데이터(200)는 인체의 관절 위치에 대한 정보를 포함할 수 있고, 관절의 x, y 좌표뿐만 아니라 깊이 정보를 나타내는 z좌표에 대한 정보를 포함할 수 있다. 다양한 행동을 인식할 수 있도록 하고 인식 정확도를 향상시키기 위하여 스켈레톤 데이터(200)는 다수의 관절에 대한 정보를 포함할 수 있고, 예를 들어 25개의 관절에 대한 위치 정보를 포함할 수 있다.
- [0047] RGB 영상(100)과 스켈레톤 데이터(200)는 모두 외부 디바이스로부터 수신될 수 있고, 이와 달리, 스켈레톤 데이터(200)는 RGB 영상(100)에서 추출될 수도 있다.
- [0048] 일 예에서, 프로세서는 키넥트(Kinect) 센서로부터 RGB 영상(100)과 스켈레톤 데이터(200)를 수신할 수 있다. 키넥트 센서는 뎁스(depth) 카메라로서 예컨대 칼라 센서와 적외선 센서를 포함할 수 있다. 이를 통해 키넥트 센서는 RGB 영상(100)을 생성할 수 있고, 영상 내 인체의 깊이 정보에 기초한 스켈레톤 데이터(200)를 생성할 수 있다. 프로세서는 키넥트 센서와 유무선으로 연결될 수 있고, 키넥트 센서로부터 RGB 영상(100)과 스켈레톤 데이터(200)를 모두 수신할 수 있다.
- [0049] 다른 예에서, 프로세서는 유무선으로 연결된 칼라 센서로부터 RGB 영상(100)을 수신하고, 수신된 RGB 영상(100)에서 스켈레톤 데이터(200)를 추출할 수 있다. 이를 위해, 프로세서는 컴퓨터 비전 알고리즘을 이용할 수 있고, 예컨대 특징점 추출(feature extraction) 알고리즘, 골격 검출(skeleton detection) 알고리즘을 이용할 수 있다. 이에 더하여 프로세서는 전술한 알고리즘을 지도 학습된(supervised learned) 신경망을 통해 구현할 수도 있다.
- [0050] 한편, 스켈레톤 데이터(200)는 RGB 영상(100)의 연속되는 프레임(110)에 대응하는 스켈레톤 시퀀스를 포함할 수 있다.
- [0051] 도 3 및 도 4를 함께 참조하면, 영상의 각 프레임(110)에서 각 관절은 그 위치에 대응하는 좌표를 가질 수 있다. 예를 들어, t번째 프레임(110)에서 천골(spine base, sacrum)의 좌표는 (x_1, y_1, z_1) 일 수 있고, 천골의 좌표는 각 프레임(110)별로 특정될 수 있다. 또한, t+a번째 프레임(110)에서 오른손 엄지(thumb right)의 좌표는 (x_{25}, y_{25}, z_{25}) 일 수 있고, 오른손 엄지의 좌표는 각 프레임(110)별로 특정될 수 있다.
- [0052] 이처럼, 모든 관절은 프레임(110)별 좌표를 가질 수 있고, 연속된 프레임(110)에 대한 연속된 좌표 정보가 스켈레톤 시퀀스일 수 있다. 이러한 스켈레톤 시퀀스는 2차원 컨볼루션 신경망(이하, 2D-CNN)에 입력될 수 있는데, 이에 대해서는 후술하도록 한다.

- [0053] 제1 스트림에서, 프로세서는 RGB 영상(100)을 3차원 컨볼루션 신경망(이하, 3D-CNN)에 입력하여(S11) 제1 분류 스코어를 획득할 수 있다(S12).
- [0054] 본 발명에서 3D-CNN은 여러 쌍의 합성곱 레이어 및 풀링(pooling) 레이어를 포함할 수 있고, 한 쌍의 합성곱 레이어 및 풀링 레이어의 출력이 다음 쌍의 합성곱 레이어 및 풀링 레이어에 입력되는 구조를 가지되, 합성곱에 이용되는 필터가 3차원으로 구성될 수 있다. 이에 따라, 후술하는 바와 같이 RGB 영상(100)의 프레임(110) 묶음을 입력 데이터로 사용하였을 때, 3D-CNN은 연속되는 프레임(110)의 시계열적인 특징을 포함하는 3차원 특징을 추출할 수 있다.
- [0055] 3D-CNN은 다양한 세부 네트워크 구조를 가질 수 있고, 예컨대 C3D(Learning Spatiotemporal Features with 3D Convolutional Networks), I3D(Two-Stream Inflated 3D ConvNet), R3D(Learning Spatio-Temporal Features with 3D Residual Networks for Action Recognition) 등의 아키텍처(architecture)를 포함할 수 있다.
- [0056] 프로세서는 RGB 영상(100)을 3차원 입력 데이터로 구성하고 이를 3D-CNN에 입력할 수 있다. 구체적으로, 프로세서는 RGB 영상(100)을 구성하는 연속된 프레임(110) 묶음을 3차원 데이터로 구성할 수 있다.
- [0057] 도 5를 참조하면, 프로세서는 RGB 영상(100)의 프레임(110) 묶음을 3D-CNN에 입력되는 3차원 입력 데이터로 가공할 수 있다. 예를 들어, RGB 영상(100)의 각 프레임(110)이 256 x 256의 크기를 갖고, RGB 영상(100)이 총 16개의 프레임(110)으로 구성되는 경우, 프로세서는 RGB 영상(100)의 프레임(110) 묶음을 256 x 256 x 16의 크기를 갖는 3차원 입력 데이터로 가공할 수 있다.
- [0058] 프로세서는 3차원 입력 데이터를 3D-CNN에 입력할 수 있고, 3D-CNN은 사전에 학습된 3차원 필터의 파라미터(가중치(weight), 바이어스(bias))에 기초하여 제1 분류 데이터를 출력할 수 있다. 구체적으로, 3D-CNN 내 마지막 풀링 레이어의 출력은 완전 연결 레이어(fully connected layer)에 입력될 수 있고, 완전 연결 레이어에서는 소프트맥스(softmax) 함수를 통해 제1 분류 데이터가 출력될 수 있다.
- [0059] 본 명세서에서 분류 데이터(제1 내지 제3 분류 데이터 포함)는 RGB 영상(100) 내 인체의 행동 클래스에 대한 확률로서, 구체적으로는 3D-CNN 내 합성곱 레이어와 풀링 레이어를 통해 추출된 특징이 각 행동 클래스에 속할 확률을 나타내는 데이터일 수 있다.
- [0060] 제2 스트림에서, 프로세서는 RGB 영상(100)의 각 프레임(110)을 2D-CNN에 입력하여(S21) 제2 분류 스코어를 획득할 수 있다(S22).
- [0061] 본 발명에서 2D-CNN은 여러 쌍의 합성곱 레이어 및 풀링 레이어를 포함할 수 있고, 한 쌍의 합성곱 레이어 및 풀링 레이어의 출력이 다음 쌍의 합성곱 레이어 및 풀링 레이어에 입력되는 구조를 가지되, 합성곱에 이용되는 필터가 2차원으로 구성될 수 있다.
- [0062] 2D-CNN은 다양한 세부 네트워크 구조를 가질 수 있고, 예컨대 GoogLeNet, ResNet(Residual Neural Network) 등의 전이학습모델(transfer learning model)을 포함할 수 있다.
- [0063] 프로세서는 RGB 영상(100)을 구성하는 각 프레임(110)별 2D 이미지를 2D-CNN에 입력할 수 있고, 2D-CNN은 사전에 학습된 2차원 필터의 파라미터에 기초하여 제2 분류 데이터를 출력할 수 있다. 구체적으로, 2D-CNN 내 마지막 풀링 레이어의 출력은 완전 연결 레이어에 입력될 수 있고, 완전 연결 레이어에서는 소프트맥스 함수를 통해 제2 분류 데이터가 출력될 수 있다.
- [0064] 한편, 행동 인식은 과거의 정보가 현재의 행동을 인식하는데 중요한 단서가 되며, 앞서 설명한 것과 같이 RGB 영상(100)은 다수의 프레임(110)으로 구성됨으로써 시계열적인 정보를 포함할 수 있다. 이러한 시계열적인 정보를 학습시키기 위하여 프로세서는 순환 신경망(Recurrent Neural Network; RNN)을 더 이용할 수 있다.
- [0065] 보다 구체적으로, 프로세서는 RGB 영상(100)의 각 프레임(110)을 2D-CNN에 입력할 수 있고, 2D-CNN을 통해 추출된 특징의 시퀀스를 RNN에 입력하여 제2 분류 스코어를 획득할 수 있다.
- [0066] 도 6을 참조하면, 프로세서는 RGB 영상(100)의 각 프레임(110)(110, V_t)를 2D-CNN에 입력하여 각 프레임(110)의 공간적 특징(p_t)을 추출할 수 있다. 이어서, 프로세서는 추출된 공간적 특징(p_t)을 RNN에 입력할 수 있고, RNN은 다수의 은닉층을 통해 과거 공간적 특징(p_t)을 재귀적으로 참조함으로써 제2 분류 스코어(q_t)를 출력할 수 있다. 제2 분류 스코어를 출력하기 위해 소프트맥스 함수가 이용될 수 있는 점은 전술한 바와 같다.
- [0067] 상기와 같이 2D-CNN과 RNN이 연결된 구조에 RGB 영상(100)의 각 프레임(110)을 입력하는 경우, RNN으로 하여금

RGB 영상(100)의 시계열적인 정보를 학습하도록 함으로써 행동 인식의 정확도를 향상시킬 수 있다.

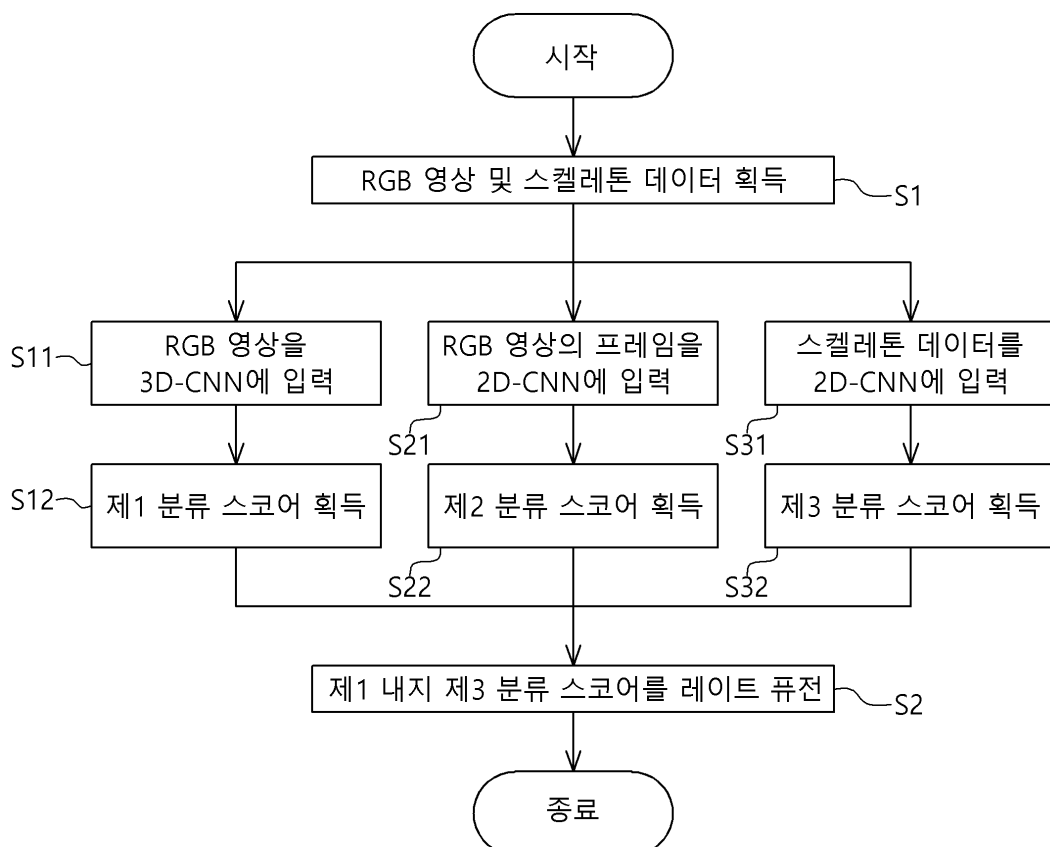
- [0068] 한편, 전통적인 RNN은 직전의 특징만을 참조하는 구조를 가지므로, 현재 행동에 영향을 주는 과거의 시점이 너무 먼 경우 RNN의 성능은 낮아질 수 있다. 이러한 성능 저하를 방지하기 위하여 프로세서는 RNN으로서 LSTM(Long Short-Term Memory) 모델을 이용할 수 있다.
- [0069] 도 7을 참조하면, 프로세서는 RGB 영상(100)의 각 프레임(110)을 2D-CNN에 입력할 수 있다. 2D-CNN은 합성곱 레이어와 풀링 레이어(예컨대, 맥스 풀링 레이어)를 통해 각 프레임(110)의 공간적 특징을 출력할 수 있고, 출력된 특징들은 시퀀스로 묶여 LSTM 모델에 입력될 수 있다. LSTM은 연속되는 레이어가 공유하는 셀 스테이트(cell state)를 통해 과거 공간적 특징을 지속적으로 참조함으로써 제2 분류 스코어를 출력할 수 있다. 제2 분류 스코어를 출력하기 위해 소프트맥스 함수가 이용될 수 있는 점은 전술한 바와 같다.
- [0070] 제3 스트림에서, 프로세서는 스켈레톤 데이터(200)를 2D-CNN에 입력하여(S31) 제3 분류 스코어를 획득할 수 있다(S32). 제3 스트림에서 이용되는 2D-CNN의 기본 구조는 제2 스트림에서 설명한 2D-CNN의 기본 구조와 동일하므로 여기서는 자세한 설명을 생략하도록 한다.
- [0071] 한편, 2D-CNN은 이미지를 입력받는 신경망인데, 도 4에서 설명한 바와 같이 스켈레톤 데이터(200)는 이미지가 아닌 좌표로 표현될 수 있다. 이에 따라, 프로세서는 스켈레톤 데이터(200)를 2D-CNN에 입력하기 위하여 이미지로 변환할 수 있다.
- [0072] 보다 구체적으로, 프로세서는 스켈레톤 데이터(200)를 RGB 이미지(300)로 변환하고, 각 프레임(110)에 대응하는 RGB 이미지(300)를 2D-CNN에 입력하여 제3 분류 스코어를 획득할 수 있다. 이 때, 프로세서는 각 프레임(110)별로 스켈레톤 데이터(200)를 RGB 이미지(300)로 변환할 수도 있고, 전체 스켈레톤 데이터(200)를 단일의 RGB 이미지(300)로 변환할 수도 있다.
- [0073] 도 8을 참조하면, 일 실시예에서 프로세서는 프레임(110)별로 특정된 각 관절의 좌표값에 기초하여 스켈레톤 데이터(200)를 RGB 이미지(300)로 변환할 수 있다.
- [0074] 도 4를 참조하여 설명한 바와 같이, 스켈레톤 데이터(200)는 프레임(110)별로 특정될 수 있다. 프로세서는 프레임(110)별 각 관절의 좌표값에 기초하여 프레임(110)별 RGB 이미지(300)를 생성할 수 있다. 구체적으로, 프로세서는 스켈레톤 데이터(200)에 포함된 프레임(110)별 관절의 3차원 좌표 값을 R, G, B 색상에 매핑하여 RGB 이미지(300)를 생성할 수 있다.
- [0075] 예를 들어, 프로세서는 관절의 x좌표 값을 정규화하여 0 내지 255 사이의 R 색상 값에 매핑할 수 있다. 마찬가지로 프로세서는 y, z좌표 값을 정규화하여 G, B 색상 값에 매핑할 수 있다. 이러한 과정을 통해 프로세서는 각 프레임(110)에 대하여 스켈레톤 데이터(200)를 RGB 이미지(300)로 변환할 수 있다. 만약, RGB 영상(100)이 16개의 프레임(110)으로 구성되는 경우 프로세서는 16개의 RGB 이미지(300)를 생성할 수 있다.
- [0076] 이어서, 프로세서는 RGB 이미지(300)를 2D-CNN에 입력하여 제3 분류 스코어를 획득할 수 있다. 2D-CNN이 합성곱 레이어와 풀링 레이어를 통해 분류 데이터를 출력하는 과정은 전술한 바 있으므로 여기서는 자세한 설명을 생략하도록 한다.
- [0077] 다만 앞서 설명한 예시에 따르면 각 프레임(110) 별로 독립된 RGB 이미지(300)가 생성됨에 따라, RGB 영상(100)이 갖는 시계열적인 정보가 소실될 수 있다. 이러한 시계열적 정보 소실을 방지하기 위하여 프로세서는 전체 스켈레톤 데이터(200)를 단일의 RGB 이미지(300)로 변환할 수 있다.
- [0078] 도 9를 참조하면, 일 실시예에서 프로세서는 스켈레톤 데이터(200)를, 스켈레톤 데이터(200)의 수와 RGB 영상(100)의 프레임(110) 수의 두 좌표축을 갖는 RGB 이미지(300)로 변환할 수 있다. 보다 구체적으로, 프로세서는 연속되는 프레임(110)에 대하여 관절(joint)별 좌표 값으로 이루어진 스켈레톤 시퀀스를, 관절별 프레임(110)으로 구성되고 이미지를 구성하는 각 픽셀이 RGB 색상으로 표현되는 RGB 이미지(300)로 변환할 수 있다.
- [0079] 도 10을 예로 들어 설명하면, RGB 영상(100)이 16개의 프레임(110)으로 구성되고 각 프레임(110)에 대한 스켈레톤 데이터(200)가 25개의 관절의 좌표를 포함하는 경우, 프로세서는 스켈레톤 시퀀스를 축변환하여 16 x 25의 크기를 갖는 이미지를 생성하되, 스켈레톤 데이터(200)를 구성하는 각 관절의 좌표를 RGB 색상에 매핑함으로써 이미지의 각 픽셀이 RGB 색상으로 표현되도록 할 수 있다.
- [0080] 프로세서는 전술한 방법에 따라 생성된 RGB 이미지(300)를 2D-CNN에 입력하여 제3 분류 스코어를 획득할 수 있다. 이 때, RGB 이미지(300)를 2D-CNN에 대응하는 입력 데이터 규격으로 변환하기 위하여, 정규화 및 선형 보간

(linear interpolation)이 이루어질 수 있음은 자명하다.

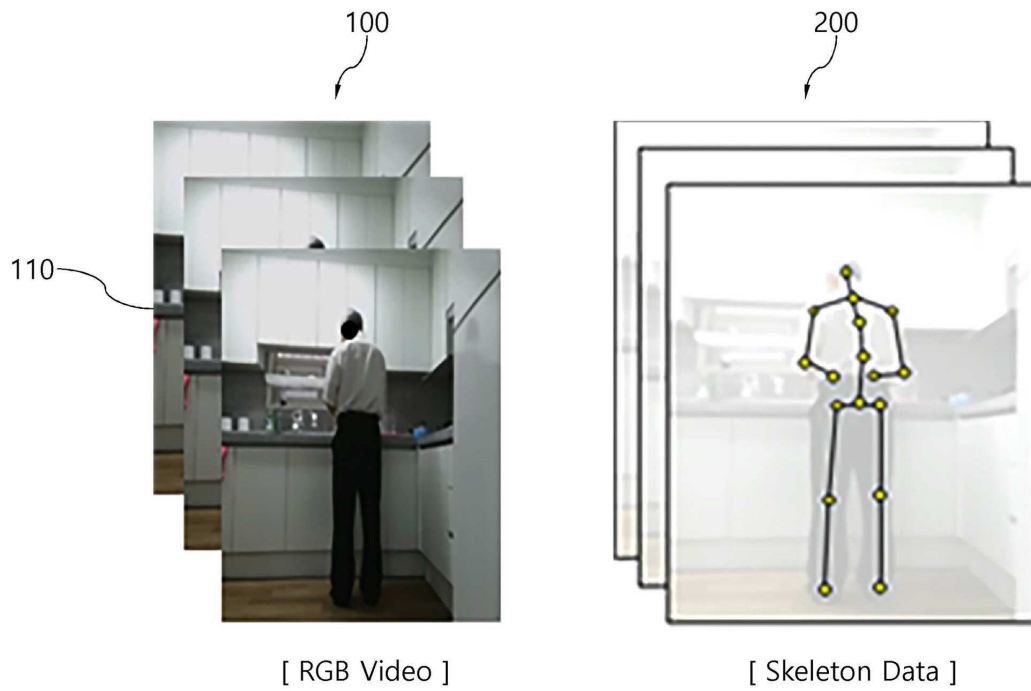
- [0081] 상기와 같이 스켈레톤 시퀀스를 단일의 RGB 이미지(300)로 변환하여 2D-CNN에 입력하는 경우, 2D-CNN으로 하여금 스켈레톤 데이터(200)가 갖는 공간적인 정보뿐만 아니라 RGB 영상(100)이 갖는 시계열적인 정보를 학습하도록 함으로써 행동 인식의 정확도를 향상시킬 수 있다.
- [0082] 전술한 제1 내지 제3 스트림에 따라 제1 내지 제3 분류 스코어가 획득되면, 프로세서는 이들을 레이트 퓨전(late-fusion)하여 행동 인식을 수행할 수 있다(S2). 구체적으로, 프로세서는 제1 내지 제3 분류 스코어를 결합하여 최종 분류 스코어를 산출할 수 있고, 이에 따라 행동 인식을 수행할 수 있다.
- [0083] 프로세서의 레이트 퓨전에는 다양한 방식이 적용될 수 있다. 예컨대, 프로세서는 제1 내지 제3 분류 스코어 행렬의 각 성분을 평균함으로써 최종 분류 스코어를 산출할 수 있다. 이와 달리, 프로세서는 제1 내지 제3 분류 스코어 행렬을 요소별로(elementwise) 곱하여 최종 분류 스코어를 산출할 수도 있다.
- [0084] 앞서 설명한 바와 같이 분류 스코어는 행동 클래스에 대한 확률을 나타내므로, 프로세서는 최종 분류 스코어 행렬 내 최대 성분에 대응하는 행동 클래스를 RGB 영상(100) 내 인체의 행동으로 결정할 수 있다.
- [0085] 전술한 본 발명은 3-스트림 앙상블 모델을 통해 서로 다른 특징을 갖는 RGB 영상(100)과 스켈레톤 데이터(200)로부터 최적의 정보를 효율적으로 추출 및 병합함으로써 신경망을 이용한 행동 인식의 신뢰성을 향상시킬 수 있다.
- [0086] 이상과 같이 본 발명에 대해서 예시한 도면을 참조로 하여 설명하였으나, 본 명세서에 개시된 실시 예와 도면에 의해 본 발명이 한정되는 것은 아니며, 본 발명의 기술사상의 범위 내에서 통상의 기술자에 의해 다양한 변형이 이루어질 수 있음은 자명하다. 아울러 앞서 본 발명의 실시 예를 설명하면서 본 발명의 구성에 따른 작용 효과를 명시적으로 기재하여 설명하지 않았을 지라도, 해당 구성에 의해 예측 가능한 효과 또한 인정되어야 함은 당연하다.

도면

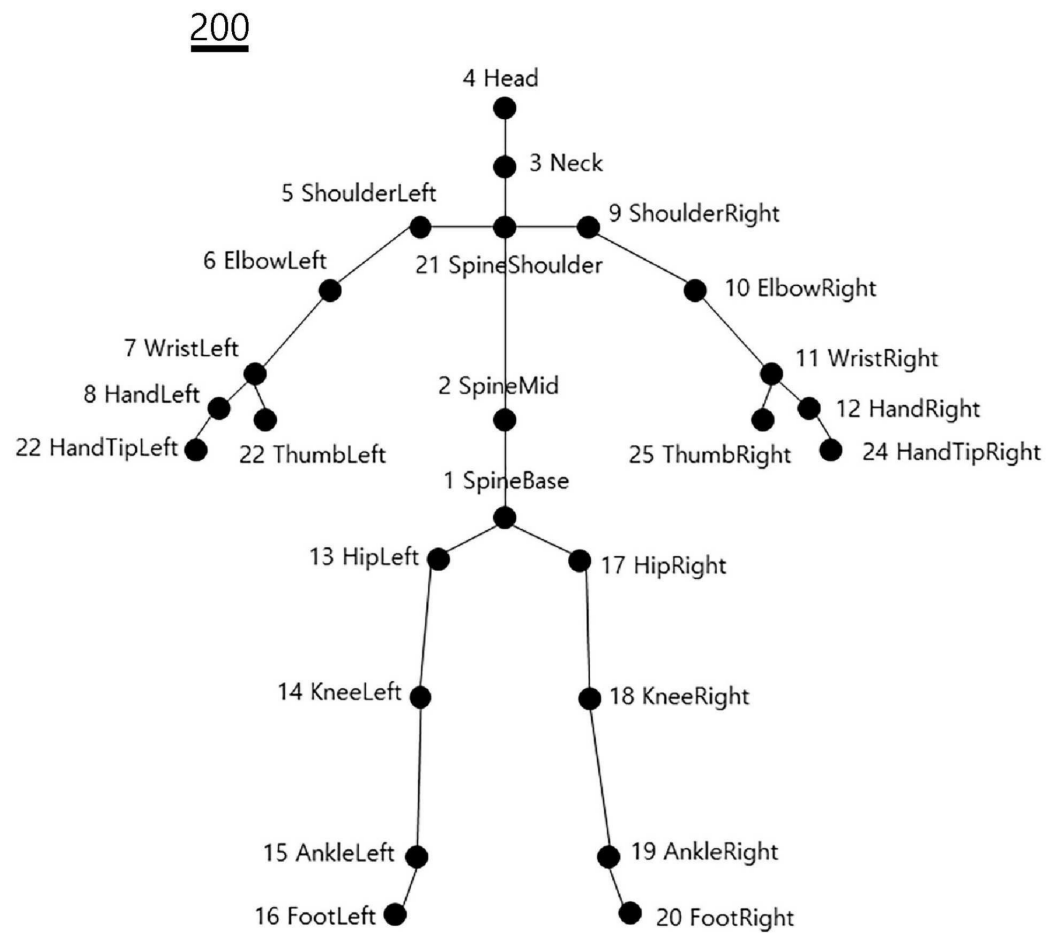
도면1



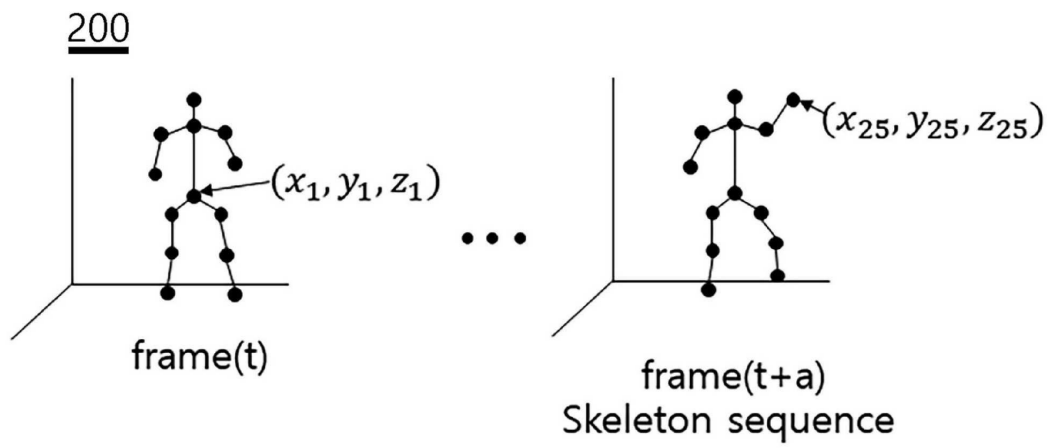
도면2



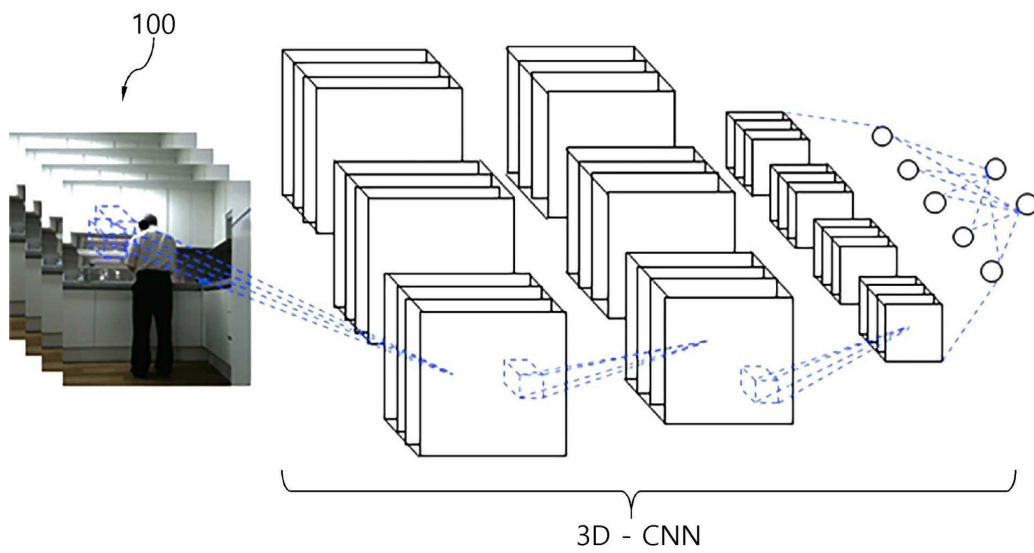
도면3



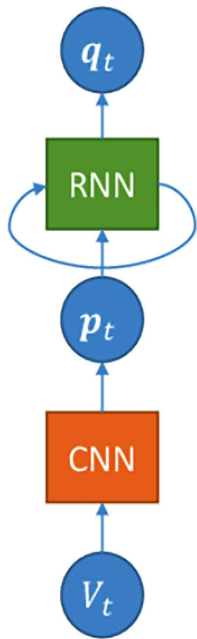
도면4



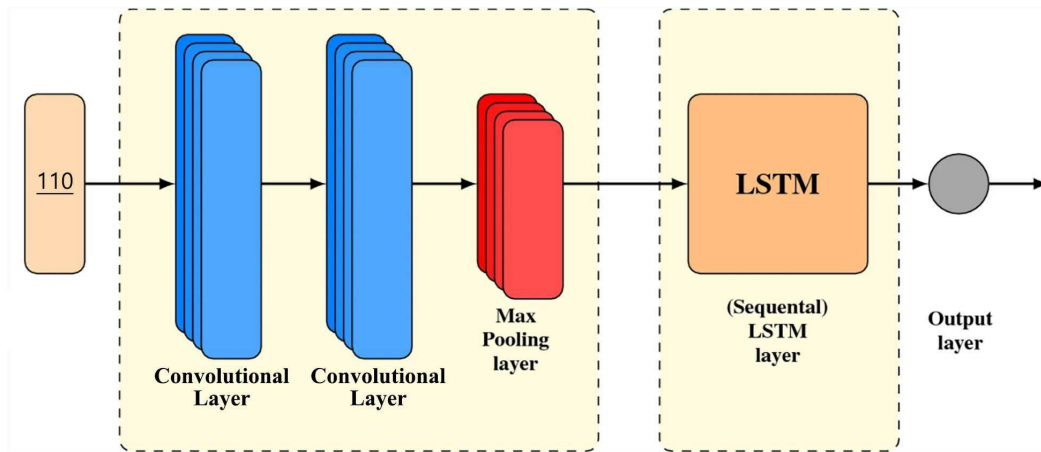
도면5



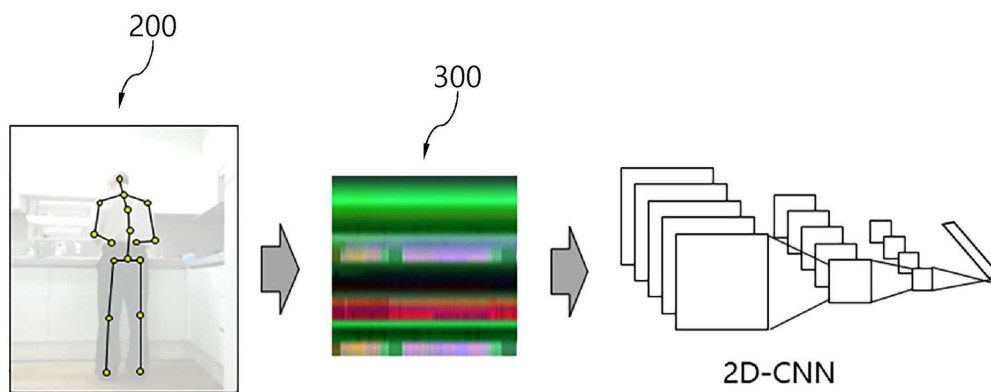
도면6



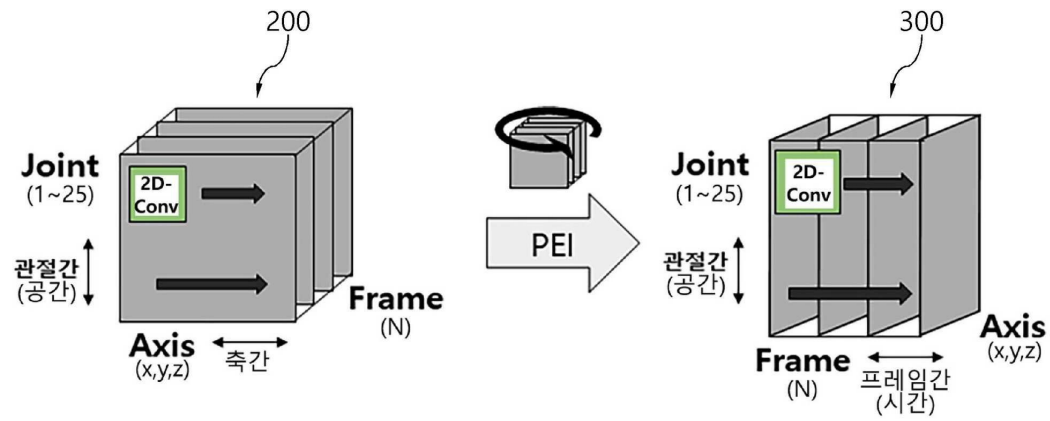
도면7



도면8



도면9



도면10

