(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0064587
(43) 공개일자 2025년05월09일

(51) 국제특허분류(Int. Cl.)
G06V 40/20 (2022.01) *G06V 10/46* (2022.01)
G06V 10/62 (2022.01) *G06V 20/40* (2022.01)
(52) CPC특허분류
G06V 40/20 (2022.01)
G06V 10/469 (2023.08)
(21) 출원번호 10-2024-0136842
(22) 출원일자 2024년10월08일
심사청구일자 2024년10월08일
(30) 우선권주장
1020230146811 2023년10월30일 대한민국(KR)

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
김상필
서울특별시 성북구 안암로 145, 우정정보관 606호 (안암동5가)
서동윤
서울시 성북구 고려대로 24가길 11, 301호
(74) 대리인
특허법인위솔

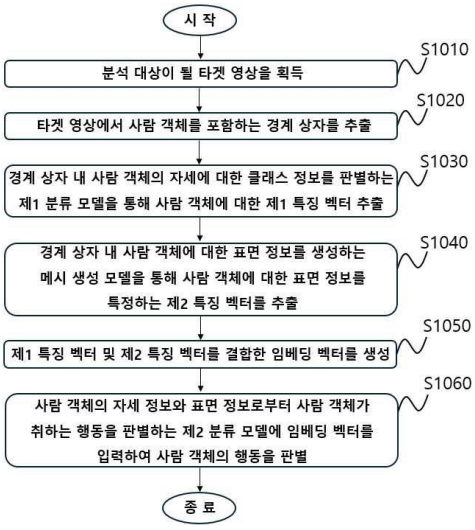
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 **사람 객체의 감지 영역에 대한 표면 평균 곡률을 활용한 행동 인식 방법 및 장치**

(57) 요 약

일 실시예에 따른 행동 인식 장치는 분석 대상이 될 타겟 영상을 획득하는 동작; 상기 타겟 영상에서 사람 객체를 포함하는 경계 상자를 추출하는 동작; 상기 경계 상자 내 사람 객체의 자세에 대한 클래스 정보를 판별하는 제1 분류 모델을 기초로 상기 사람 객체에 대한 제1 특징 벡터를 추출하는 동작; 상기 경계 상자 내 사람 객체에 대한 표면 정보를 생성하는 메시 생성 모델을 기초로 상기 사람 객체에 대한 표면 정보를 특정하는 제2 특징 벡터를 추출하는 동작; 상기 제1 특징 벡터 및 제2 특징 벡터를 결합한 임베딩 벡터를 생성하는 동작; 및 사람 객체의 자세 정보와 표면 정보로부터 사람 객체가 취하는 행동을 판별하는 제2 분류 모델에 상기 임베딩 벡터를 입력하여 상기 사람 객체의 행동을 판별하는 동작을 포함할 수 있다.

대 표 도 - 도2



(52) CPC특허분류
G06V 10/62 (2023.08)
G06V 20/49 (2022.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 1711193973
과제번호 2019-0-00079-005
부처명 과학기술정보통신부
과제관리(전문)기관명 정보통신기획평가원
연구사업명 정보통신방송혁신인재양성
연구과제명 인공지능대학원지원(고려대학교)
과제수행기관명 고려대학교산학협력단
연구기간 2023.01.01 ~ 2023.12.31
이 발명을 지원한 국가연구개발사업
과제고유번호 1345370620
과제번호 LINC3.0-2023-1
부처명 교육부
과제관리(전문)기관명 한국연구재단
연구사업명 산학협력고도화지원
연구과제명 3단계산학연협력선도대학육성사업(LINC3.0)(0.5)
과제수행기관명 고려대학교(서울)
연구기간 2023.03.01 ~ 2024.02.29
이 발명을 지원한 국가연구개발사업
과제고유번호 2370000057
과제번호 RS-2024-00345025
부처명 문화체육관광부
과제관리(전문)기관명 한국콘텐츠진흥원
연구사업명 저작권기술 글로벌 인재양성
연구과제명 생성형 AI 저작권 관리 및 보호 기술 개발을 위한 국제공동연구 및 글로벌 인재 양성
과제수행기관명 고려대학교 산학협력단
연구기간 2024.04.01 ~ 2027.12.31
이 발명을 지원한 국가연구개발사업
과제고유번호 2370000058
과제번호 RS-2024-00348469
부처명 문화체육관광부
과제관리(전문)기관명 한국콘텐츠진흥원
연구사업명 신기술 융합 저작권 기술개발
연구과제명 생성형 AI 3D 콘텐츠 저작권 보호를 위한 뉴럴 워터마크 기술 연구
과제수행기관명 고려대학교 산학협력단
연구기간 2024.04.01 ~ 2025.12.31

명 세 서

청구범위

청구항 1

프로세서에 의해 동작하는 행동 인식 장치가 수행하는 방법에 있어서,

분석 대상이 될 타겟 영상을 획득하는 동작;

상기 타겟 영상에서 사람 객체를 포함하는 경계 상자를 추출하는 동작;

상기 경계 상자 내 사람 객체의 자세에 대한 클래스 정보를 판별하는 제1 분류 모델을 기초로 상기 사람 객체에 대한 제1 특징 벡터를 추출하는 동작;

상기 경계 상자 내 사람 객체에 대한 표면 정보를 생성하는 메시 생성 모델을 기초로 상기 사람 객체에 대한 표면 정보를 특정하는 제2 특징 벡터를 추출하는 동작;

상기 제1 특징 벡터 및 제2 특징 벡터를 결합한 임베딩 벡터를 생성하는 동작; 및

사람 객체의 자세 정보와 표면 정보로부터 사람 객체가 취하는 행동을 판별하는 제2 분류 모델에 상기 임베딩 벡터를 입력하여 상기 사람 객체의 행동을 판별하는 동작을 포함하는,

방법.

청구항 2

제1항에 있어서,

상기 제2 특징 벡터를 추출하는 동작은

메시 생성 모델 중 ROMP 모델을 이용하여 상기 사람 객체가 차지하는 영역의 3차원 메시 정보를 복원하는 동작을 포함하는,

방법.

청구항 3

제2항에 있어서,

상기 메시 정보는

상기 사람 객체가 차지하는 영역에 대한 3차원 정보를 나타내는 복수의 정점 및 간선으로 구성된,

방법.

청구항 4

제3항에 있어서,

상기 제2 특징 벡터를 추출하는 동작은

상기 복수의 정점 중 각 정점의 평균 곡률을 계산하는 동작; 및

상기 각 정점의 평균 곡률에 대한 정보를 벡터화한 제2 특징 벡터를 생성하는 동작을 포함하는,

방법.

청구항 5

제4항에 있어서,

상기 평균 곡률을 계산하는 동작은

상기 복수의 정점 중 어느 하나인 제1 정점과, 상기 제1 정점의 1-링 주변부(1-ring neighborhood)에 해당하는 각 정점 간의 곡률값을 계산하는 동작; 및

상기 곡률값 중 최대 곡률과 최소 곡률의 평균을 상기 제1 정점의 평균 곡률로 계산하는 동작을 포함하는,

방법.

청구항 6

제1항에 있어서,

상기 임베딩 벡터를 생성하는 동작은

Concatenation, Addition, Averaging, Gate network 중 적어도 하나의 기법을 기초로 상기 제1 특징 벡터 및 제2 특징 벡터를 결합하여 임베딩 벡터를 생성하는 동작을 포함하는,

방법.

청구항 7

제1항에 있어서,

상기 제2 분류 모델은

시계열 데이터를 처리하는 LSTM 모델을 기반으로 사람 객체의 자세 정보와 표면 정보로부터 구성된 임베딩 벡터와 사람 객체의 행동을 특정하는 정답 클래스 간의 상관 관계에 대한 파라미터를 학습하도록 지도 학습된,

방법.

청구항 8

제7항에 있어서,

상기 제2 분류 모델은

상기 타겟 영상 중의 복수의 프레임 각각으로부터 추출된 제1 특징 벡터와 제2 특징 벡터의 결합으로 생성된 복수의 임베딩 벡터를 입력받아 상기 타겟 영상에 포함된 사람 객체의 행동을 판별하는,

방법.

청구항 9

제8항에 있어서,

상기 복수의 프레임은

상기 타겟 영상의 프레임 중 시계열적인 순서에 따라 추출된 일련의 프레임을 포함하고,

상기 제2 분류 모델은

상기 복수의 임베딩 벡터를 각 임베딩 벡터의 프레임의 시계열적 순서에 따라 입력 받아, 상기 복수의 임베딩 벡터의 시계열적 순서를 반영하여 타겟 영상에 포함된 사람 객체의 행동을 판별하는 것이 특징인,

방법.

청구항 10

명령어를 포함하는 메모리; 및

상기 명령어를 기초로 소정의 동작을 수행하는 프로세서를 포함하고,

상기 프로세서의 동작은,

분석 대상이 될 타겟 영상을 획득하는 동작;

상기 타겟 영상에서 사람 객체를 포함하는 경계 상자를 추출하는 동작;

상기 경계 상자 내 사람 객체의 자세에 대한 클래스 정보를 판별하는 제1 분류 모델을 기초로 상기 사람 객체에 대한 제1 특징 벡터를 추출하는 동작;

상기 경계 상자 내 사람 객체에 대한 표면 정보를 생성하는 메시 생성 모델을 기초로 상기 사람 객체에 대한 표면 정보를 특정하는 제2 특징 벡터를 추출하는 동작;

상기 제1 특징 벡터 및 제2 특징 벡터를 결합한 임베딩 벡터를 생성하는 동작; 및

사람 객체의 자세 정보와 표면 정보로부터 사람 객체가 취하는 행동을 판별하는 제2 분류 모델에 상기 임베딩 벡터를 입력하여 상기 사람 객체의 행동을 판별하는 동작을 포함하는,

행동 인식 장치.

발명의 설명

기 술 분 야

[0001] 본 발명은 인공지능 및 컴퓨터 비전 기술에 관한 것으로서, 사람 객체의 자세 정보 및 표면 정보를 융합한 임베딩 벡터의 분석을 통해 객체의 행동을 정밀하게 인식하는 기술에 관한 것이다.

배 경 기 술

[0002] 인공지능 기술의 발전에 따라 컴퓨터 비전을 활용한 행동 인식 기술은 영상을 분석하여 사람의 행동을 인식하고 추정하는 중요한 방법론으로 자리 잡고 있다. 행동 인식 기술은 보안 시스템에서 침입자나 이상 행동을 감지하고, 헬스케어에서는 환자의 상태를 모니터링하며, 스포츠 분석에서는 선수의 동작을 평가하고 향상시키는 데 활용된다. 또한, 스마트 환경에서 행동 인식 기술은 사용자의 움직임을 파악해 편리함을 증대시키는 데 기여할 수 있다.

[0003] 한편, 기존의 행동 인식 기술은 객체 감지와 관절 인식 기술을 중심으로 발전해왔다. 객체 감지 기술은 영상에서 사람 또는 물체를 경계상자(bounding box)로 감지하여 객체의 위치와 크기를 추출한다. 객체 감지 기술은 영상 속에서 특정 객체가 어디에 위치하고, 어떻게 이동하는지를 파악하는 데 강점이 있다.

[0004] 또한, 관절 인식 기술은 사람의 신체에서 주요 관절 좌표를 추출하여 이를 기반으로 골격 모델을 형성한다. 이 골격 모델을 통해 사람의 자세나 움직임을 분석할 수 있는데, 이는 사람이 수행하는 특정 동작을 추정하는 데 중요한 기초 데이터를 제공할 수 있다. 예를 들면, 스포츠 선수의 자세나 환자의 움직임을 분석하여 건강 상태를 진단할 수 있다.

[0005] 그러나 이러한 기존의 기술들은 각각의 한계점을 지니고 있다. 객체 감지 기술은 주로 객체의 존재 여부와 경계 상자 좌표에 초점을 맞추기 때문에, 사람의 세부적인 행동을 파악하기 위한 정보가 부족하다. 예를 들어, 손동작과 같은 미세한 신체 움직임을 인식하는 데에는 한계가 있다. 또한, 객체의 위치 정보만으로 복잡한 행동을 구분하기에는 어려움이 있으며, 특히 다수의 객체가 상호작용하는 상황에서는 더더욱 한계를 드러낸다.

[0006] 관절 인식 기술 역시 사람의 특정 관절 좌표에만 의존하므로, 관절이 명확하게 드러나지 않는 영상에서는 오탐지의 가능성이 높아지고, 세부적인 동작을 정확하게 인식하기 어렵다. 특히, 사람과 물체 간의 상호작용을 필요로 하는 행동을 추정할 때, 관절 정보만으로는 충분한 의미를 도출하기 어렵다. 예를 들어, 사람이 물체를 잡거나 특정 동작을 수행하는 상황에서 관절 정보는 중요한 단서를 제공하지만, 그 외에 필요한 추가적인 행동 정보는 제공하지 못한다는 한계가 있다.

[0007] 이러한 기존 기술의 한계는 더 높은 수준의 행동 인식을 요구하는 응용 분야에서 기술적 성능을 제약하는 요소로 작용하고 있다. 이로 인해, 복잡한 동작이나 다양한 상호작용을 포함하는 행동을 정확하게 인식하기 위해서는 추가적인 데이터의 활용과 고도화된 분석 기법이 필요한 실정이다.

선행기술문헌

특허문헌

[0008] (특허문헌 0001) 대한민국 등록특허공보 제10-2238085호

발명의 내용

해결하려는 과제

[0009] 본 발명이 해결하고자 하는 과제는 사람 객체의 자세 정보와, 객체의 표면을 메시화하여 도출한 평균 곡률을 활용하여 더욱 정밀한 행동 인식을 구현하는 것이다. 표면 곡률은 사람의 자세와 움직임을 세밀하게 분석할 수 있는 중요한 정보를 제공하며, 이를 통해 기존의 관절 위치나 경계상자 정보만으로는 부족했던 세부적인 행동까지 추정할 수 있게 한다. 이는 기존 기술로는 인식하기 어려운 미세한 동작까지도 보다 정확하게 인식할 수 있도록 하여, 복잡한 행동 분석에 있어 큰 이점을 제공한다.

[0010] 또한, 본 발명은 시계열 네트워크를 도입하여 시간의 흐름에 따른 행동 변화를 분석하는 기술을 제공하는 것을 목표로 한다. 단일 프레임 분석만으로는 정확한 행동 예측이 어려운 경우가 많기 때문에, 본 발명은 다중 프레임을 활용하여 시간에 따른 행동 변화를 포착하고, 이를 통해 동적이고 복잡한 행동을 더욱 정확하게 추정하고자 한다. 이를 통해 사람의 연속적인 움직임을 반영하여, 정적인 정보에 국한되지 않고 동적인 정보까지 포함한 종합적인 행동 인식이 가능하도록 한다.

[0011] 따라서, 본 발명은 사람 객체의 자세 정보와 메시 정보를 통합한 임베딩 벡터, 그리고 시계열 네트워크를 활용한 행동 인식 기술을 제안하여 종래의 행동 인식 기술을 개선하고자 한다. 이에 따라, 본 발명의 기술은 기존 기술에 비해 더 높은 정확도와 세밀함을 제공하여, 보안, 헬스케어, 스포츠 분석 등 다양한 분야에서 복잡하고 다양한 행동을 보다 정확하게 인식하는 기술적 기반을 제공하는 것을 목표로 한다. 이러한 개선은 다양한 산업 및 응용 분야에서 행동 인식의 정확성을 크게 향상시키며, 복잡한 행동 패턴을 더욱 효과적으로 분석할 수 있는 유연하고 강력한 시스템을 구현하는 데 기여할 것이다.

[0012] 한편, 본 발명의 기술적 과제들은 이상에서 언급한 기술적 과제들로 제한되지 않으며, 언급되지 않은 또 다른 기술적 과제들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

[0013] 일 실시예에 따른 프로세서에 의해 동작하는 행동 인식 장치가 수행하는 방법에 있어서, 분석 대상이 될 타겟 영상을 획득하는 동작; 상기 타겟 영상에서 사람 객체를 포함하는 경계 상자를 추출하는 동작; 상기 경계 상자 내 사람 객체의 자세에 대한 클래스 정보를 판별하는 제1 분류 모델을 기초로 상기 사람 객체에 대한 제1 특징 벡터를 추출하는 동작; 상기 경계 상자 내 사람 객체에 대한 표면 정보를 생성하는 메시 생성 모델을 기초로 상기 사람 객체에 대한 표면 정보를 특정하는 제2 특징 벡터를 추출하는 동작; 상기 제1 특징 벡터 및 제2 특징 벡터를 결합한 임베딩 벡터를 생성하는 동작; 및 사람 객체의 자세 정보와 표면 정보로부터 사람 객체가 취하는 행동을 판별하는 제2 분류 모델에 상기 임베딩 벡터를 입력하여 상기 사람 객체의 행동을 판별하는 동작을 포함할 수 있다.

[0014] 또한, 상기 제2 특징 벡터를 추출하는 동작은 메시 생성 모델 중 ROMP 모델을 이용하여 상기 사람 객체가 차지하는 영역의 3차원 메시 정보를 복원하는 동작을 포함할 수 있다.

[0015] 또한, 상기 메시 정보는 상기 사람 객체가 차지하는 영역에 대한 3차원 정보를 나타내는 복수의 정점 및 간선으로 구성될 수 있다.

[0016] 또한, 상기 제2 특징 벡터를 추출하는 동작은 상기 복수의 정점 중 각 정점의 평균 곡률을 계산하는 동작; 및 상기 각 정점의 평균 곡률에 대한 정보를 벡터화한 제2 특징 벡터를 생성하는 동작을 포함할 수 있다.

[0017] 또한, 상기 평균 곡률을 계산하는 동작은 상기 복수의 정점 중 어느 하나인 제1 정점과, 상기 제1 정점의 1-링 주변부(1-ring neighborhood)에 해당하는 각 정점 간의 곡률값을 계산하는 동작; 및 상기 곡률값 중 최대 곡률과 최소 곡률의 평균을 상기 제1 정점의 평균 곡률로 계산하는 동작을 포함할 수 있다.

[0018] 또한, 상기 임베딩 벡터를 생성하는 동작은 Concatenation, Addition, Averaging, Gate network 중 적어도 하나의 기법을 기초로 상기 제1 특징 벡터 및 제2 특징 벡터를 결합하여 임베딩 벡터를 생성하는 동작을 포함할 수 있다.

[0019] 또한, 상기 제2 분류 모델은 시계열 데이터를 처리하는 LSTM 모델을 기반으로 사람 객체의 자세 정보와 표면 정보로부터 구성된 임베딩 벡터와 사람 객체의 행동을 특정하는 정답 클래스 간의 상관 관계에 대한 파라미터를 학습하도록 지도 학습될 수 있다.

[0020] 또한, 상기 제2 분류 모델은 상기 타겟 영상 중의 복수의 프레임 각각으로부터 추출된 제1 특징 벡터와 제2 특징 벡터의 결합으로 생성된 복수의 임베딩 벡터를 입력받아 상기 타겟 영상에 포함된 사람 객체의 행동을 판별할 수 있다.

[0021] 또한, 상기 복수의 프레임은 상기 타겟 영상의 프레임 중 시계열적인 순서에 따라 추출된 일련의 프레임을 포함하고, 상기 제2 분류 모델은 상기 복수의 임베딩 벡터를 각 임베딩 벡터의 프레임의 시계열적 순서에 따라 입력받아, 상기 복수의 임베딩 벡터의 시계열적 순서를 반영하여 타겟 영상에 포함된 사람 객체의 행동을 판별할 수 있다.

[0022] 일 실시예에 따른 행동 인식 장치는 명령어를 포함하는 메모리; 및 상기 명령어를 기초로 소정의 동작을 수행하는 프로세서를 포함하고, 상기 프로세서의 동작은 분석 대상이 될 타겟 영상을 획득하는 동작; 상기 타겟 영상에서 사람 객체를 포함하는 경계 상자를 추출하는 동작; 상기 경계 상자 내 사람 객체의 자세에 대한 클래스 정보를 판별하는 제1 분류 모델을 기초로 상기 사람 객체에 대한 제1 특징 벡터를 추출하는 동작; 상기 경계 상자 내 사람 객체에 대한 표면 정보를 생성하는 메시 생성 모델을 기초로 상기 사람 객체에 대한 표면 정보를 특정하는 제2 특징 벡터를 추출하는 동작; 상기 제1 특징 벡터 및 제2 특징 벡터를 결합한 임베딩 벡터를 생성하는 동작; 및 사람 객체의 자세 정보와 표면 정보로부터 사람 객체가 취하는 행동을 판별하는 제2 분류 모델에 상기 임베딩 벡터를 입력하여 상기 사람 객체의 행동을 판별하는 동작을 포함할 수 있다.

발명의 효과

[0023] 본 발명은 사람 객체의 자세 정보와 표면 정보를 통합한 임베딩 벡터, 그리고 시계열 네트워크를 활용하여, 보다 정밀하고 정확한 행동 인식을 가능하게 한다.

[0024] 구체적으로, 본 발명은 객체의 메시 정보에서 도출한 곡률 정보를 활용함으로써, 기존의 경계상자나 관절 정보만으로는 파악하기 어려운 세부적인 자세 변화와 동작을 정밀하게 인식할 수 있도록 한다. 예를 들어, 사람의 신체 표면에서 발생하는 작은 곡률 변화를 분석하여 보다 세밀한 행동 인식이 가능해지며, 이는 복잡한 동작이나 다양한 상황에서 매우 유용하게 작용할 수 있다.

[0025] 또한, 본 발명은 시계열 네트워크를 도입하여 시간의 흐름에 따른 행동 변화를 다중 프레임으로 분석함으로써, 단일 프레임 분석보다 더 높은 정확도의 행동 예측을 제공한다. 이는 연속적인 동작과 복잡한 행동을 인식하는 데 있어 큰 장점을 제공하며, 정적인 장면뿐만 아니라 동적인 상황에서도 행동을 효과적으로 추정할 수 있게 한다. 특히, 시간의 흐름에 따른 동작 변화를 분석함으로써 행동 인식의 신뢰성이 높아지고, 자연스러운 행동 예측을 실현할 수 있다.

[0026] 이처럼, 본 발명은 경계상자에 대한 객체의 자세 정보, 메시 기반 곡률 정보, 그리고 시계열 네트워크를 연계함으로써, 다양한 환경에서 사람의 행동을 정확하게 인식할 수 있는 유연성을 제공한다. 이로 인해 고정된 이미지나 동영상뿐만 아니라 실시간 카메라 영상에서도 높은 정확도를 유지할 수 있다. 이러한 유연성은 실시간 모니터링 시스템이나 산업 현장 등 다양한 실용적인 응용 분야에 효과적으로 적용될 수 있다.

[0027] 따라서, 본 발명은 기존 기술에 비해 행동 인식의 정확도와 정밀도를 크게 향상시키며, 인공지능 기반 행동 인식 기술의 발전에 기여할 수 있다. 이는 헬스케어, 스포츠 분석, 스마트 환경 등 다양한 응용 분야에서 유용하게 활용될 수 있으며, 더욱 높은 수준의 행동 인식 성능을 제공하여 응용 범위를 넓히는 데 중요한 역할을 할 것이다.

[0028] 한편, 본 발명에 의한 효과는 이상에서 언급한 것들로 제한되지 않으며, 언급되지 않은 또 다른 기술적 효과들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

[0029] 도 1은 일 실시예에 따른 행동 인식 장치의 구성도이다.

도 2는 일 실시예에 따른 행동 인식 장치가 수행하는 동작의 흐름도이다.

도 3은 일 실시예에 따른 행동 인식 장치가 수행하는 동작의 전체 과정을 개념화한 예시도이다.

도 4는 일 실시예에 따른 행동 인식 장치가 YOLO 네트워크 구조 기반의 객체 검출 모델을 활용하는 예시도이다.

도 5는 일 실시예에 따른 행동 인식 장치가 RESNET 네트워크 구조 기반의 분류 모델을 활용하는 예시도이다.

도 6 내지 도 8은 일 실시예에 따른 제1 분류 모델이 판별하는 사람 객체의 자세에 대한 클래스의 예시도이다.

도 9는 일 실시예에 따른 행동 인식 장치가 경계 상자 내 사람 객체에 대한 표면 정보를 생성하는 동작의 예시도이다.

도 10은 일 실시예에 따른 행동 인식 장치가 제2 분류 모델로서 LSTM 모델을 활용하여 사람 객체가 취하는 행동 클래스를 도출하는 동작의 예시도이다.

발명을 실시하기 위한 구체적인 내용

[0030] 본 발명의 목적과 기술적 구성 및 그에 따른 작용 효과에 관한 자세한 사항은 본 발명의 명세서에 첨부된 도면에 의거한 이하의 상세한 설명에 의해 보다 명확하게 이해될 것이다. 첨부된 도면을 참조하여 본 발명에 따른 실시예를 상세하게 설명한다.

[0031] 본 명세서에서 개시되는 실시예들은 본 발명의 범위를 한정하는 것으로 해석되거나 이용되지 않아야 할 것이다. 이 분야의 통상의 기술자에게 본 명세서의 실시예를 포함한 설명은 다양한 응용을 갖는다는 것이 당연하다. 따라서, 본 발명의 상세한 설명에 기재된 임의의 실시예들은 본 발명을 보다 잘 설명하기 위한 예시적인 것이며 본 발명의 범위가 실시예들로 한정되는 것을 의도하지 않는다.

[0032] 도면에 표시되고 아래에 설명되는 기능 블록들은 가능한 구현의 예들일 뿐이다. 다른 구현들에서는 상세한 설명의 사상 및 범위를 벗어나지 않는 범위에서 다른 기능 블록들이 사용될 수 있다. 또한, 본 발명의 하나 이상의 기능 블록이 개별 블록들로 표시되지만, 본 발명의 기능 블록들 중 하나 이상은 동일 기능을 실행하는 다양한 하드웨어 및 소프트웨어 구성들의 조합일 수 있다.

[0033] 또한, 어떤 구성요소들을 포함한다는 표현은 "개방형"의 표현으로서 해당 구성요소들이 존재하는 것을 단순히 지칭할 뿐이며, 추가적인 구성요소들을 배제하는 것으로 이해되어서는 안 된다.

[0034] 나아가 어떤 구성요소가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급될 때에는, 그 다른 구성요소에 직접적으로 연결 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 한다.

[0035] 이하, 본 발명의 다양한 실시예가 첨부된 도면을 참조하여 기재된다. 그러나, 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 실시예의 다양한 변경(modification), 균등물(equivalent), 및/또는 대체물(alternative)을 포함하는 것으로 이해되어야 한다.

[0036] 본 발명은 사람 객체의 행동 인식을 위한 기술을 제공하는 것으로, 객체의 자세 정보와 표면 메시에서 추출된 곡률 정보를 결합하여 정밀한 행동 인식을 수행하는 기술을 제안한다.

[0037] 이하, 도 1 내지 도 10과 함께, 본 발명의 실시예를 구현하는 행동 인식 장치(100)의 구성과, 행동 인식 장치(100)가 수행하는 동작들을 살펴본다.

[0038] 도 1은 일 실시예에 따른 행동 인식 장치(100)(이하, '장치(100)'로 지칭)의 구성도이다.

[0039] 도 1을 참조하면, 일 실시예에 따른 장치(100)는 각각 메모리(110), 프로세서(120), 입출력 인터페이스(130) 및 통신 인터페이스(140)를 포함할 수 있다.

[0040] 메모리(110)는 외부 장치(100)로부터 획득한 데이터 또는 스스로 생성한 데이터를 저장할 수 있다. 메모리(110)는 프로세서(120)의 동작을 수행시킬 수 있는 명령어들을 저장할 수 있다. 예를 들어, 메모리(110)는 본 문서의 실시예에 따라 분석 대상이 될 타겟 영상, 타겟 영상의 분석에 사용할 하나 이상의 객체 검출 모델(ex. YOLO, R-CNN, SSD 등), 분류 모델(ex. RESNET, ALEXNET, Inception, LSTM 등), 메시 생성 모델(ex. ROMP, SMPL, PIFu, DEEPLABCUT3D, HRM 등)을 저장할 수 있다.

[0041] 프로세서(120)는 전반적인 동작을 제어하는 연산 장치(100)이다. 프로세서(120)는 메모리(110)에 저장된 명령어들을 실행할 수 있다. 본 문서의 실시예에 따른 장치(100)의 동작은 프로세서(120)에 의해 수행되는 동작으로 이해될 수 있다.

[0042] 입출력 인터페이스(130)는 정보를 입력하거나 출력하는 하드웨어 인터페이스 또는 소프트웨어 인터페이스를 포함할 수 있다.

[0043] 통신 인터페이스(140)는 통신망을 통해 정보를 송수신 할 수 있게 한다. 이를 위해, 통신 인터페이스(140)는 무

선 통신모듈 또는 유선 통신모듈을 포함할 수 있다.

[0044] 장치(100)는 프로세서(120)를 통해 연산을 수행하고 네트워크를 통해 정보를 송수신할 수 있는 다양한 형태의 장치(100)로 구현될 수 있다. 예를 들면, 서버, 컴퓨터 장치, 휴대용 통신 장치, 스마트 폰, 휴대용 멀티미디어 장치, 노트북, 태블릿 PC 등의 형태로 구현될 수 있으나, 이러한 예시에 한정되는 것은 아니다.

[0045] 도 2는 일 실시예에 따른 장치(100)가 수행하는 동작의 흐름도이다. 도 3은 일 실시예에 따른 장치(100)가 수행하는 동작의 전체 과정을 개념화한 예시도이다. 도 2 및 도 3의 실시예에 따른 장치(100)의 동작은 프로세서(120)에 의해 수행되는 동작으로 이해될 수 있다.

[0046] 도 2 및 도 3에 개시된 각 단계는 본 발명의 목적을 달성함에 있어서 바람직한 실시예일 뿐이며, 필요에 따라 일부 단계가 추가 또는 삭제될 수 있음은 물론이고, 어느 한 단계가 다른 단계에 포함되어 수행될 수도 있다. 도 2 및 도 3에 개시된 각 동작의 순서는 이해의 편의를 위해 배치된 순서일 뿐, 이러한 순서가 시계열적인 순서로 한정되는 것이 아니며, 설계자의 선택에 따라 순서가 다르게 변경되어 동작될 수 있다.

[0047] 도 2 및 도 3를 참조하면, S1010 단계에서, 장치(100)는 분석 대상이 될 타겟 영상을 획득할 수 있다.

[0048] 일 예로, 장치(100)는 다양한 외부 장치로부터 사람 객체가 포함된 영상 데이터를 확보할 수 있다. 타겟 영상은 동영상 파일이나 실시간 스트리밍 영상 등 다양한 형태로 제공될 수 있다. 예를 들어, 타겟 영상은 CCTV 영상, 스마트폰 카메라로 촬영된 영상, 또는 웹캠 영상 등 다양한 형식의 입력 영상 데이터를 포함할 수 있다. 가령, 타겟 영상은 범죄 예방을 위해 실시간 촬영 중인 길거리 CCTV 영상이 될 수 있다.

[0049] 장치(100)는 이러한 타겟 영상을 입력 데이터로 활용하여, 후술할 단계를 통해 타겟 영상 속에 포함된 사람 객체를 분석하여 사람 객체의 행동을 인식하게 된다.

[0050] S1020 단계에서, 장치(100)는 타겟 영상에서 사람 객체를 포함하는 경계 상자(bounding box)를 추출할 수 있다. 장치(100)는 타겟 영상을 처리하여 사람 객체를 식별하고, 사람 객체를 둘러싸는 경계 상자를 설정할 수 있다. 예를 들어, 장치(100)는 길거리 CCTV 영상에서 보행자, 자전거를 타고 있는 사람 등을 탐지하여 각 사람 객체 주위에 경계 상자를 생성할 수 있다.

[0051] 이를 위해, 장치(100)는 객체 탐지 알고리즘을 활용하여 타겟 영상에서 사람 객체를 포함하는 경계 상자를 추출할 수 있다. 예를 들어, 장치(100)는 YOLO(You Only Look Once), R-CNN(Region-based Convolutional Neural Networks), SSD(Single Shot MultiBox Detector) 등과 같은 딥러닝 기반의 객체 탐지 알고리즘을 활용하여 영상의 각 프레임에서 사람 객체의 위치를 정확하게 추출할 수 있다.

[0052] 도 4는 일 실시예에 따른 장치(100)가 YOLO 네트워크 구조 기반의 객체 검출 모델을 활용하는 예시도이다.

[0053] 도 4를 참조하면, 일 실시예에 따른 YOLO는 객체 검출 모델로서 입력된 이미지 또는 영상에서 다양한 객체를 탐지하고, 각각의 객체에 대한 경계 상자와, 특정 경계 상자 속의 객체가 속하는 클래스에 대한 확률을 예측할 수 있다.

[0054] 일 예로, 장치(100)는 416x416 크기의 타겟 영상을 네트워크의 입력으로 사용할 수 있다. 입력 이미지는 다양한 크기의 객체를 포함할 수 있으며, YOLO 네트워크는 타겟 영상을 다음과 같이 처리하여 객체를 검출할 수 있다.

[0055] 첫번째 단계로, Convolution 레이어는 3x3 크기의 필터 32개를 사용하여 Convolution 연산을 수행하여, 타겟 영상의 공간적 특징을 추출하게 된다. 또한, MaxPooling 레이어는 2x2 크기의 필터를 사용하여 공간 크기를 절반으로 줄인다. 이 과정을 통해 208x208 크기의 특성 맵이 생성될 수 있다.

[0056] 두번째 단계로, Convolution 레이어는 3x3 크기의 필터 64개를 사용하여 Convolution 연산을 수행하고, MaxPooling 레이어를 통해 다시 공간 크기를 절반으로 줄여 104x104 크기의 특성 맵을 생성할 수 있다.

[0057] 세번째 단계로, Convolution 레이어는 3x3 크기의 필터 128개를 사용하여 Convolution 연산을 수행하고, MaxPooling 레이어를 통해 52x52 크기의 특성 맵이 생성된다.

[0058] 네번째 단계로, Convolution 레이어는 3x3 크기의 필터 256개를 사용하여 Convolution 연산을 수행하고, MaxPooling 레이어를 통해 26x26 크기의 특성 맵이 생성된다.

[0059] 다섯번째 단계로, Convolution 레이어는 3x3 크기의 필터 512개를 사용하여 Convolution 연산을 수행하고, MaxPooling 레이어를 통해 13x13 크기의 특성 맵이 생성된다. 이때 Reorg 레이어는 26x26 크기의 특성 맵을 13x13 크기로 재구성(reorganize)하여, 특성 맵의 깊이를 두 배로 늘릴 수 있고, 이는 네트워크가 더 많은 정보

를 유지할 수 있도록 하여, 객체 검출의 정확도를 향상시킬 수 있다.

[0060] 마지막 단계로, 여러 Convolution 레이어를 통해 최종적으로 13x13x1024 크기의 특성 맵이 생성되며, 이 단계에서는 3x3 크기의 필터와 1x1 크기의 필터를 번갈아 사용하여 연산을 수행할 수 있다.

[0061] 이에 따라, 출력 레이어에서, 최종 출력은 13x13x125 크기의 특성 맵이 출력될 수 있다. 여기서 125는 각 그리드 셀마다 5개의 경계 상자와 각 경계 상자에 대해 25개의 출력(4개의 경계 상자 좌표, 1개의 객체 존재 확률, 20개의 클래스 확률)을 포함할 수 있다.

[0062] 장치(100)는 이러한 객체 검출 모델을 통해, 사람 객체의 위치뿐만 아니라 크기, 모양 등의 다양한 정보를 포함하도록 경계 상자를 생성할 수 있다. 또한, 장치(100)는 S1020 단계에서 추출된 경계 상자 중 사람 객체에 해당하는 영역을 크롭핑하여 다음 S1030 및 S1040 단계의 입력으로 사용할 수 있다. 즉, S1020 단계에서 추출된 경계 상자 중 사람 객체에 해당하는 영역은 이후의 단계에서 사람 객체의 자세 정보와 메시 정보를 추출하는 기초 데이터로 활용될 수 있다.

[0063] S1030 단계에서, 장치(100)는 경계 상자 내의 사람 객체의 자세에 대한 클래스 정보를 판별하는 제1 분류 모델을 사용하여 경계 상자에 포함된 사람 객체에 대한 제1 특징 벡터를 추출할 수 있다. 이를 위해, 장치(100)는 딥러닝 기반의 제1 분류 모델을 활용하여 경계 상자 내의 사람 객체가 취하는 자세에 대한 클래스를 판별할 수 있다. 제1 분류 모델은 경계 상자 내의 픽셀 정보를 분석하여, 사람 객체가 취하는 자세에 대한 클래스를 판별할 수 있다. 예를 들어, 제1 분류 모델은 RESNET, ALEXNET, Inception 등과 같은 딥러닝 기반의 분류 모델을 포함할 수 있다.

[0064] 도 5는 일 실시예에 따른 장치(100)가 RESNET 네트워크 구조 기반의 분류 모델을 활용하는 예시도이다.

[0065] 도 5를 참조하면, ResNet은 딥러닝 모델 중 하나로 주로 이미지의 클래스 분류에 사용되며, 깊은 네트워크에서도 효과적으로 학습할 수 있도록 설계된 모델이다. ResNet의 주요 특징은 Residual Block을 사용하여 네트워크의 깊이가 깊어져도 기울기 소실 문제를 극복할 수 있다는 것이다.

[0066] ResNet의 입력 단계에서는 3x3 크기의 Convolution 레이어로 시작된다. 이는 입력 이미지의 특징을 추출하는 초기 단계로, 이미지의 저수준(low-level) 특징을 학습한다. 이때 각 Residual Block은 3x3 크기의 Convolution 레이어로 구성되며, 여러 개의 Residual Block이 연속으로 쌓여 있다. Residual Block 내부에서는 입력이 다음 레이어의 출력에 더해지는(skip connection) 구조로 되어 있다. 이는 네트워크가 더 깊어지더라도 학습이 원활하게 이루어지도록 한다. 이러한 Residual Block은 네트워크가 더 많은 계층을 학습할 수 있도록 도와주며, 기울기 소실 문제를 방지한다.

[0067] ResNet의 중간 단계에서는 여러 개의 3x3 Convolution 레이어가 연속으로 적용된다. 이 과정에서 네트워크는 이미지의 중간 수준(mid-level) 특징을 학습한다. 여기서도 Residual Block이 적용되며, 각 블록의 출력은 다음 블록의 입력에 더해진다. 다음의 깊은 단계에서는 더 많은 3x3 Convolution 레이어가 연속으로 적용된다. 이 과정에서 네트워크는 이미지의 고수준(high-level) 특징을 학습한다. 깊은 단계에서도 Residual Block이 적용되며, 네트워크의 깊이가 깊어짐에 따라 더 복잡한 특징을 학습할 수 있다.

[0068] ResNet의 출력 단계에서는 Softmax 레이어가 적용되어 각 클래스에 대한 확률을 계산한다. Softmax 레이어는 이미지가 각 클래스에 속할 확률을 출력한다.

[0069] 본 문서의 실시예는 Softmax 레이어가 구분할 클래스를 다음 도 6 내지 도 8의 예시와 같이 학습시킬 수 있다.

[0070] 도 6 내지 도 8은 일 실시예에 따른 제1 분류 모델이 판별하는 사람 객체의 자세에 대한 클래스의 예시도이다.

[0071] 도 6은 사람의 다양한 자세를 나타내는 클래스의 예시를 보여준다. 도 6에 따르면, 제1 분류 모델은 사람의 자세를 포괄적인 카테고리를 우선적으로 구분하고, 해당 카테고리의 세부 분류를 분류하도록 학습될 수 있다.

[0072] 도 6 및 도 7을 함께 참조하면, 서 있는 자세(310)는 직립 자세(311), 한쪽 다리를 굽히고 서 있는 자세(312), 양쪽 다리를 굽히고 서 있는 자세(313)로 세분화될 수 있다.

[0073] 도 6 및 도 7을 함께 참조하면, 앉아 있는 자세(320)는 의자에 등을 기대고 앉아 있는 자세(321), 다리를 굽힌 채로 앉아 있는 자세(322), 다리를 굽히고 무릎을 가슴에 대고 앉아 있는 자세(323), 다리를 굽히지 않고 뻗은 채로 앉아 있는 자세(324)로 세분화될 수 있다.

[0074] 도 6 및 도 8을 함께 참조하면, 누워 있는 자세(330)는 등을 바닥에 대고 다리를 뻗어 누운 자세(331), 등을 바

닥에 대고 다리를 접어 누운 자세(332), 측면으로 몸을 대고 다리를 뻗어 누운 자세(333), 측면으로 몸을 대고 다리를 접어 누운 자세(334), 엎드리고 다리를 뻗어 누운 자세(335), 엎드리고 다리를 접어 누운 자세(336)로 세분화될 수 있다.

[0075] 도 6 및 도 8을 함께 참조하면, 운동 자세(340)는 요가에서 사용되는 자세(341), 다리를 구부린 점프 자세(342), 다이빙 준비 자세(343), 달리기 자세(344)로 세분화될 수 있다.

[0076] 도 6 및 도 8을 함께 참조하면, 팔에 대한 자세(350)는 양 팔을 내린 자세(351), 한쪽 팔을 몸 쪽으로 접은 자세(352), 양쪽 팔을 몸 쪽으로 접은 자세(353), 한쪽 팔만 정면으로 뻗은 자세(354), 양쪽 팔을 정면으로 뻗은 자세(355), 한쪽 팔만 하늘 방향으로 뻗은 자세(356), 양쪽 팔을 하늘 방향으로 뻗은 자세(357)로 세분화될 수 있다.

[0077] 이처럼, 제1 분류 모델은 ResNet 구조의 신경망에 대한 딥러닝 학습을 기초로, 사람 객체의 경계 상자와 사람 객체의 자세를 특정하는 정답 클래스 간의 상관 관계에 대한 파라미터가 지도 학습되어, 도 2의 동작 전에 미리 생성될 수 있다.

[0078] 한편, 제1 분류 모델은 다중 출력 분류 네트워크 방식으로 학습될 수 있으며, 이 경우 하나의 경계 상자에 대해 복수의 자세 클래스를 판별할 수 있도록 설계될 수 있다. 예를 들어, 길거리에서 우산을 쓰고 서 있는 사람을 포함하는 경계 상자를 입력하였다고 가정하면, 제1 분류 모델은 일 직선으로 서 있는 자세(311), 한쪽 팔만 하늘 방향으로 뻗은 자세(356)과 같이 둘 이상의 클래스를 출력할 수 있다.

[0079] 장치(100)는 이러한 제1 분류 모델에 경계 상자를 입력하였을 때, 제1 분류 모델 내부의 레이어 중 어느 하나의 레이어로부터 출력되는 특징 벡터를 제1 특징 벡터로 추출할 수 있다. 예를 들어, 제1 분류 모델의 레이어 중 Softmax 레이어에 입력될 특징 벡터는 경계 상자에 포함된 사람 객체의 자세 클래스를 판별할 수 있는 정보가 특정된 벡터로 구성되어 있다고 볼 수 있다. 따라서, 장치(100)는 제1 분류 모델 내부의 특징 벡터 중 Softmax 레이어에 입력될 특징 벡터를 제1 특징 벡터로서 추출하여, 이후 단계에서 활용할 수 있다.

[0080] S1040 단계에서, 장치(100)는 경계 상자 내 사람 객체에 대한 표면 정보를 특정하는 제2 특징 벡터를 추출할 수 있다. 예를 들어, 표면 정보는 3차원 메시 정보를 포함할 수 있다.

[0081] 도 9는 일 실시예에 따른 장치(100)가 경계 상자 내 사람 객체에 대한 표면 정보를 생성하는 동작의 예시도이다.

[0082] 도 9를 참조하면, 장치(100)는 앞선 S1010 및 S1020 단계로부터 추출된 경계 상자 내 사람 객체에 대한 표면 정보를 생성하고, 이러한 표면 정보를 특정하는 제2 특징 벡터를 추출할 수 있다. 장치(100)는 경계 상자 내 사람 객체에 대한 영역의 2차원 이미지로부터 3차원 메시 정보를 추출하는 메시 생성 모델을 이용할 수 있다.

[0083] 여기서, 메시 생성 모델(Mesh Generation Model)이란 딥러닝 학습을 기반으로 2차원 이미지를 3차원 메시 정보로 복원하는 모델이다. 예를 들어, 장치(100)는 메시 생성 모델 중 하나인 ROMP(Regression of Multiple 3D People) 모델을 이용하여, S1010 및 S1020 단계에 따라 추출된 사람 객체 영역에 대한 2차원 이미지를 3차원 메시 정보로 복원할 수 있다. 이때 메시 정보는 사람 객체가 차지하는 영역에 대한 3차원 정보를 특정하는 복수의 정점(vertex) 및 간선(edge)을 포함할 수 있다.

[0084] 일 예로, 장치(100)는 사람 객체가 차지하는 영역의 메시 정보에 포함된 각 정점의 평균 곡률에 대한 정보를 기초로, 사람 객체의 표면 정보를 특정하는 제2 특징 벡터를 생성할 수 있다.

[0085] 장치(100)는 다음의 실시예에 따라 메시 정보에 포함된 각 정점의 평균 곡률을 계산할 수 있다.

[0086] 먼저, 장치(100)는 사람 객체가 차지하는 영역 중 어느 하나인 제1 정점과, 제1 정점의 1-링 주변부에 해당하는 각 정점 간의 곡률값을 계산할 수 있다. 여기서, 1-링 주변부(1-ring neighborhood)란 메시 구조에서 특정 정점과 하나의 간선로 직접 연결된 이웃 정점들을 의미한다.

[0087] 즉, 제1 정점은 평균 곡률을 계산하고자 하는 특정 중심점이며, 1-링 주변부는 제1 정점과 직접 연결된 인접 정점들로 구성된다. 제1 정점과 1-링 주변부의 정점들은 메시 구조에서 서로 인접하여 삼각형 면을 구성함에 따라 국소적인 곡면을 형성한다. 이때 장치(100)는 이러한 국소적인 곡면을 2차원 평면에 근사하여 해당 곡면의 국소적인 형태를 추정할 수 있다. 예를 들어, 장치(100)는 국소적인 곡면에 대한 1차 미분값 및 2차 미분값을 기초로 곡면의 기울기와 곡률값을 도출할 수 있다. 이때 장치(100)는 제1 정점의 1-링 주변부에 해당하는 각 정점 간에 계산된 곡률값 중 최대 곡률과 최소 곡률의 평균(=주곡률의 평균)을 제1 정점의 평균 곡률로 계산할 수 있

다.

[0088] 이처럼, 장치(100)는 사람 객체가 차지하는 영역에 포함된 모든 정점에 대해 위 동작을 수행하여, 사람 객체 영역의 모든 정점에 대한 평균 곡률값을 도출하고, 모든 정점의 평균 곡률값을 벡터화하여 사람 객체의 표면 정보를 특정하는 제2 특징 벡터를 생성할 수 있다.

[0089] S1050 단계에서, 장치(100)는 제1 특징 벡터 및 제2 특징 벡터를 결합한 임베딩 벡터를 생성할 수 있다. 예를 들어, 장치(100)는 Concatenation, Addition, Averaging, Gate network 등의 기법을 통해 앞서 추출된 제1 특징 벡터와 제2 특징 벡터를 통합하여 하나의 임베딩 벡터를 생성할 수 있다. 이에 따라 생성된 임베딩 벡터는 사람 객체의 자세 정보와 표면 정보의 관계를 효과적으로 표현할 수 있도록 설계된 벡터로서, 제1 벡터 및 제2 벡터의 정보가 결합되어 행동 인식에 필요한 중요한 특징을 포함하게 된다.

[0090] S1060 단계에서, 장치(100)는 사람 객체의 자세 정보와 표면 정보로부터 사람 객체가 취하는 행동을 판별하는 제2 분류 모델에 임베딩 벡터를 입력하여 타겟 영상에 포함된 사람의 행동을 인식할 수 있다. 이를 위해, 장치(100)는 딥러닝 기반의 제2 분류 모델을 활용하여 사람 객체의 자세 정보와 표면 정보의 상호작용에 따른 사람의 행동에 대한 클래스를 판별할 수 있다. 예를 들어, 제2 분류 모델은 LSTM, RESNET, ALEXNET, Inception 등과 같은 딥러닝 기반의 분류 모델을 포함할 수 있다.

[0091] 일 예로, 제2 분류 모델은 시계열 데이터를 처리하는 LSTM(Long Short-Term Memory) 모델 기반의 딥러닝 분류 모델을 포함할 수 있다. LSTM 모델은 장기 의존성 문제를 해결하기 위해 설계된 딥러닝 모델로서, 장기 기억을 보존하면서도 단기적인 정보를 효과적으로 처리할 수 있는 특징을 가지고 있어, 시간적 순서가 중요한 시계열 데이터를 처리하는 데 적합하다는 특징을 가진다.

[0092] 이를 통해, 장치(100)는 LSTM 모델을 활용해 임베딩 벡터 내의 정보를 분석하여, 사람이 취하는 행동의 클래스를 판별할 수 있다. 이때 LSTM 모델은 사람 객체의 자세 정보와 표면 정보로부터 구성된 임베딩 벡터와 사람 객체의 행동을 특정하는 정답 클래스 간의 상관 관계에 대해 기 학습된 파라미터를 가질 수 있다. 즉, S1060 단계에서 제2 분류 모델은 S1010 내지 S1040 단계와 같은 방식에 따라 기 생성된 임베딩 벡터와, 해당 임베딩 벡터에 대한 정답 클래스를 기초로 미리 지도 학습된 상태이다.

[0093] 도 10은 일 실시예에 따른 장치(100)가 제2 분류 모델로서 LSTM 모델을 활용하여 사람 객체가 취하는 행동 클래스를 도출하는 동작의 예시도이다.

[0094] 도 10을 참조하면, 제2 분류 모델은 타겟 영상 중의 복수의 프레임 각각으로부터 추출된 제1 특징 벡터와 제2 특징 벡터의 결합으로 생성된 복수의 임베딩 벡터를 입력받아 타겟 영상에 포함된 사람 객체의 행동을 판별할 수 있다. 이때 복수의 프레임은 타겟 영상의 프레임 중 시계열적인 순서에 따라 추출된 일련의 프레임을 포함할 수 있다. 즉, 제2 분류 모델은 복수의 임베딩 벡터를 각 임베딩 벡터의 프레임의 시계열적 순서에 따라 입력 받아, 복수의 임베딩 벡터의 시계열적 순서를 반영하여 타겟 영상에 포함된 사람 객체의 행동을 판별할 수 있다.

[0095] 상술한 실시예에 따르면, 본 발명은 사람 객체의 자세 정보와 표면 정보를 통합한 임베딩 벡터, 그리고 시계열 네트워크를 활용하여, 보다 정밀하고 정확한 행동 인식을 가능하게 한다.

[0096] 구체적으로, 본 발명은 객체의 메시 정보에서 도출한 곡률 정보를 활용함으로써, 기존의 경계상자나 관절 정보만으로는 파악하기 어려운 세부적인 자세 변화와 동작을 정밀하게 인식할 수 있도록 한다. 예를 들어, 사람의 신체 표면에서 발생하는 작은 곡률 변화를 분석하여 보다 세밀한 행동 인식이 가능해지며, 이는 복잡한 동작이나 다양한 상황에서 매우 유용하게 작용할 수 있다.

[0097] 또한, 본 발명은 시계열 네트워크를 도입하여 시간의 흐름에 따른 행동 변화를 다중 프레임으로 분석함으로써, 단일 프레임 분석보다 더 높은 정확도의 행동 예측을 제공한다. 이는 연속적인 동작과 복잡한 행동을 인식하는 데 있어 큰 장점을 제공하며, 정적인 장면뿐만 아니라 동적인 상황에서도 행동을 효과적으로 추정할 수 있게 한다. 특히, 시간의 흐름에 따른 동작 변화를 분석함으로써 행동 인식의 신뢰성이 높아지고, 자연스러운 행동 예측을 실현할 수 있다.

[0098] 이처럼, 본 발명은 경계상자에 대한 객체의 자세 정보, 메시 기반 곡률 정보, 그리고 시계열 네트워크를 연계함으로써, 다양한 환경에서 사람의 행동을 정확하게 인식할 수 있는 유연성을 제공한다. 이로 인해 고정된 이미지나 동영상뿐만 아니라 실시간 카메라 영상에서도 높은 정확도를 유지할 수 있다. 이러한 유연성은 실시간 모니터링 시스템이나 산업 현장 등 다양한 실용적인 응용 분야에 효과적으로 적용될 수 있다.

[0099] 따라서, 본 발명은 기존 기술에 비해 행동 인식의 정확도와 정밀도를 크게 향상시키며, 인공지능 기반 행동 인

식 기술의 발전에 기여할 수 있다. 이는 헬스케어, 스포츠 분석, 스마트 환경 등 다양한 응용 분야에서 유용하게 활용될 수 있으며, 더욱 높은 수준의 행동 인식 성능을 제공하여 응용 범위를 넓히는 데 중요한 역할을 할 것이다.

[0100] 본 문서의 다양한 실시예들 및 이에 사용된 용어들은 본 문서에 기재된 기술적 특징들을 특정한 실시예들로 한정하려는 것이 아니며, 해당 실시예의 다양한 변경, 균등물, 또는 대체물을 포함하는 것으로 이해되어야 한다. 도면의 설명과 관련하여, 유사한 또는 관련된 구성요소에 대해서는 유사한 참조 부호가 사용될 수 있다. 아이템에 대응하는 명사의 단수 형은 관련된 문맥상 명백하게 다르게 지시하지 않는 한, 아이템 한 개 또는 복수 개를 포함할 수 있다.

[0101] 본 문서에서, "A 또는 B", "A 및 B 중 적어도 하나", "A 또는 B 중 적어도 하나,""A, B 또는 C," "A, B 및 C 중 적어도 하나,"및 "A, B, 또는 C 중 적어도 하나"와 같은 문구들 각각은 그 문구들 중 해당하는 문구에 함께 나열된 항목들의 모든 가능한 조합을 포함할 수 있다. " 1", "제2", 또는 "첫째" 또는 "둘째"와 같은 용어들은 단순히 해당 구성요소를 다른 해당 구성요소와 구분하기 위해 사용될 수 있으며, 해당 구성요소들을 다른 측면(예: 중요성 또는 순서)에서 한정하지 않는다. 어떤(예: 제1) 구성요소가 다른(예: 제2) 구성요소에, "기능적으로" 또는 "통신적으로"라는 용어와 함께 또는 이런 용어 없이, "커플드" 또는 "커넥티드"라고 언급된 경우, 그것은 어떤 구성요소가 다른 구성요소에 직접적으로(예: 유선으로), 무선으로, 또는 제3 구성요소를 통하여 연결될 수 있다는 것을 의미한다.

[0102] 본 문서에서 사용된 용어 "모듈"은 하드웨어, 소프트웨어 또는 펌웨어로 구현된 유닛을 포함할 수 있으며, 예를 들면, 로직, 논리 블록, 부품, 또는 회로 등의 용어와 상호 호환적으로 사용될 수 있다. 모듈은, 일체로 구성된 부품 또는 하나 또는 그 이상의 기능을 수행하는, 부품의 최소 단위 또는 그 일부가 될 수 있다. 예를 들면, 일 실시예에 따르면, 모듈은 ASIC(application-specific integrated circuit)의 형태로 구현될 수 있다.

[0103] 본 문서의 다양한 실시예들은 기기(예: 전자 장치)에 의해 읽을 수 있는 저장 매체(예: 메모리)에 저장된 하나 이상의 명령어들을 포함하는 소프트웨어(예: 프로그램)로서 구현될 수 있다. 저장 매체는 RAM(random access memory), 메모리 버퍼, 하드 드라이브, 데이터베이스, EPROM(erasable programmable read-only memory), EEPROM(electrically erasable read-only memory), ROM(read-only memory) 및/또는 등등을 포함할 수 있다.

[0104] 또한, 본 문서의 실시예들의 프로세서는, 저장 매체로부터 저장된 하나 이상의 명령어들 중 적어도 하나의 명령을 호출하고, 그것을 실행할 수 있다. 이것은 기기가 호출된 적어도 하나의 명령어에 따라 적어도 하나의 기능을 수행하도록 운영되는 것을 가능하게 한다. 이러한 하나 이상의 명령어들은 컴파일러에 의해 생성된 코드 또는 인터프리터에 의해 실행될 수 있는 코드를 포함할 수 있다. 프로세서는 범용 프로세서, FPGA(Field Programmable Gate Array), ASIC(Application Specific Integrated Circuit), DSP(Digital Signal Processor) 및/또는 등등 일 수 있다.

[0105] 기기로 읽을 수 있는 저장매체는, 비일시적(non-transitory) 저장매체의 형태로 제공될 수 있다. 여기서, '비일시적'은 저장매체가 실재(tangible)하는 장치이고, 신호(예: 전자기파)를 포함하지 않는다는 것을 의미할 뿐이며, 이 용어는 데이터가 저장매체에 반영구적으로 저장되는 경우와 임시적으로 저장되는 경우를 구분하지 않는다.

[0106] 본 문서에 개시된 다양한 실시예들에 따른 방법은 컴퓨터 프로그램 제품(computer program product)에 포함되어 제공될 수 있다. 컴퓨터 프로그램 제품은 상품으로서 판매자 및 구매자 간에 거래될 수 있다. 컴퓨터 프로그램 제품은 기기로 읽을 수 있는 저장 매체(예: compact disc read only memory (CD-ROM))의 형태로 배포되거나, 또는 어플리케이션 스토어(예: 플레이 스토어)를 통해 또는 두 개의 사용자 장치들(예: 스마트폰들) 간에 직접, 온라인으로 배포(예: 다운로드 또는 업로드)될 수 있다. 온라인 배포의 경우에, 컴퓨터 프로그램 제품의 적어도 일부는 제조사의 서버, 어플리케이션 스토어의 서버, 또는 서버의 메모리와 같은 기기로 읽을 수 있는 저장 매체에 적어도 일시 저장되거나, 임시적으로 생성될 수 있다.

[0107] 다양한 실시예들에 따르면, 기술한 구성요소들의 각각의 구성요소(예: 모듈 또는 프로그램)는 단수 또는 복수의 개체를 포함할 수 있다. 다양한 실시예들에 따르면, 전술한 해당 구성요소들 중 하나 이상의 구성요소들 또는 동작들이 생략되거나, 또는 하나 이상의 다른 구성요소들 또는 동작들이 추가될 수 있다. 대체적으로 또는 추가적으로, 복수의 구성요소들(예: 모듈 또는 프로그램)은 하나의 구성요소로 통합될 수 있다. 이런 경우, 통합된 구성요소는 복수의 구성요소들 각각의 구성요소의 하나 이상의 기능들을 통합 이전에 복수의 구성요소들 중 해당 구성요소에 의해 수행되는 것과 동일 또는 유사하게 수행할 수 있다. 다양한 실시예들에 따르면, 모듈, 프로

그램 또는 다른 구성요소에 의해 수행되는 동작들은 순차적으로, 병렬적으로, 반복적으로, 또는 휴리스틱하게 실행되거나, 동작들 중 하나 이상이 다른 순서로 실행되거나, 생략되거나, 또는 하나 이상의 다른 동작들이 추가될 수 있다.

부호의 설명

[0108] 100: 장치

110: 메모리

120: 프로세서

130: 입출력 인터페이스

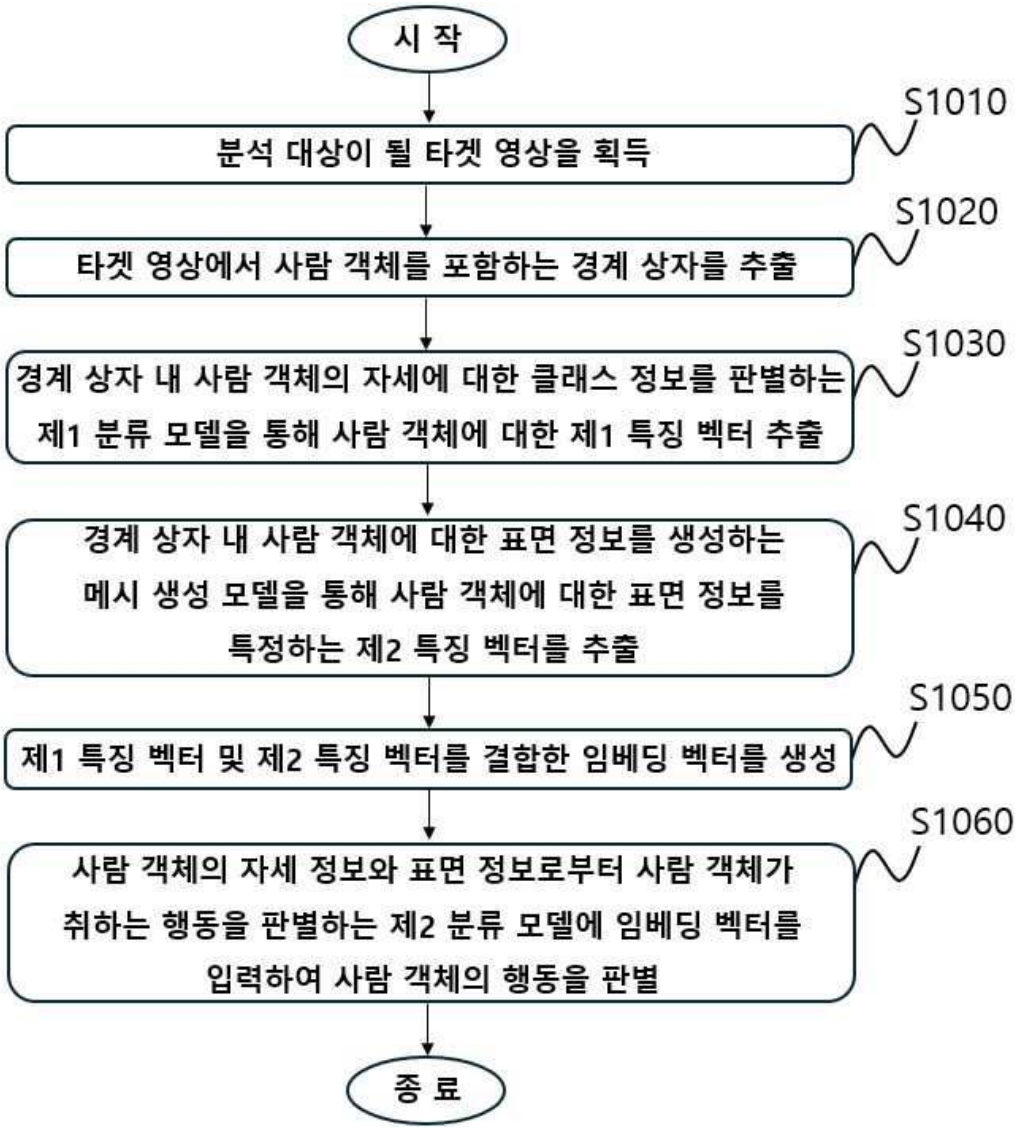
140: 통신 인터페이스

도면

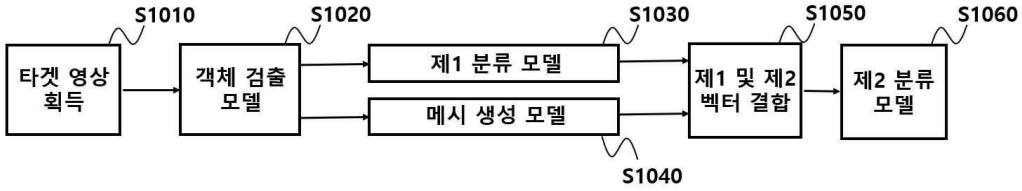
도면1



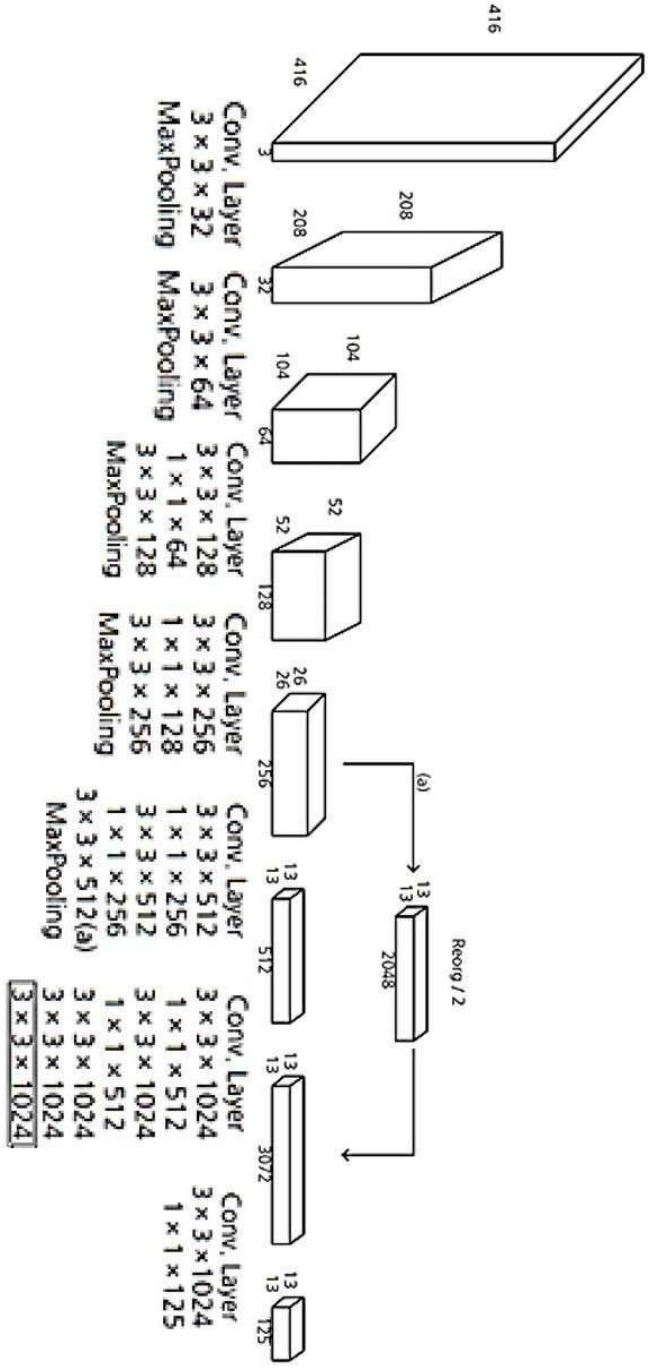
도면2


시 작
S1010
분석 대상이 될 타겟 영상을 획득
S1020
타겟 영상에서 사람 객체를 포함하는 경계 상자를 추출
S1030
경계 상자 내 사람 객체의 자세에 대한 클래스 정보를 판별하는 제1 분류 모델을 통해 사람 객체에 대한 제1 특징 벡터 추출
S1040
경계 상자 내 사람 객체에 대한 표면 정보를 생성하는 메시 생성 모델을 통해 사람 객체에 대한 표면 정보를 특정하는 제2 특징 벡터를 추출
S1050
제1 특징 벡터 및 제2 특징 벡터를 결합한 임베딩 벡터를 생성
S1060
사람 객체의 자세 정보와 표면 정보로부터 사람 객체가 취하는 행동을 판별하는 제2 분류 모델에 임베딩 벡터를 입력하여 사람 객체의 행동을 판별
종 료

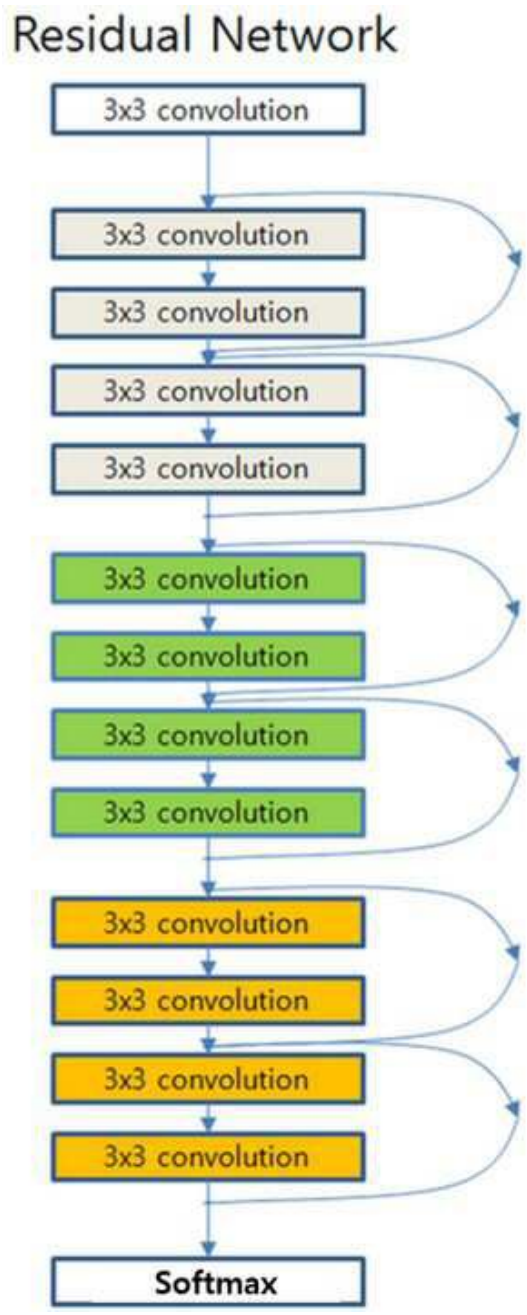
도면3


S1010
타겟 영상 획득
S1020
객체 검출 모델
제1 분류 모델
메시 생성 모델
S1030
S1040
S1050
제1 및 제2 벡터 결합
S1060
제2 분류 모델

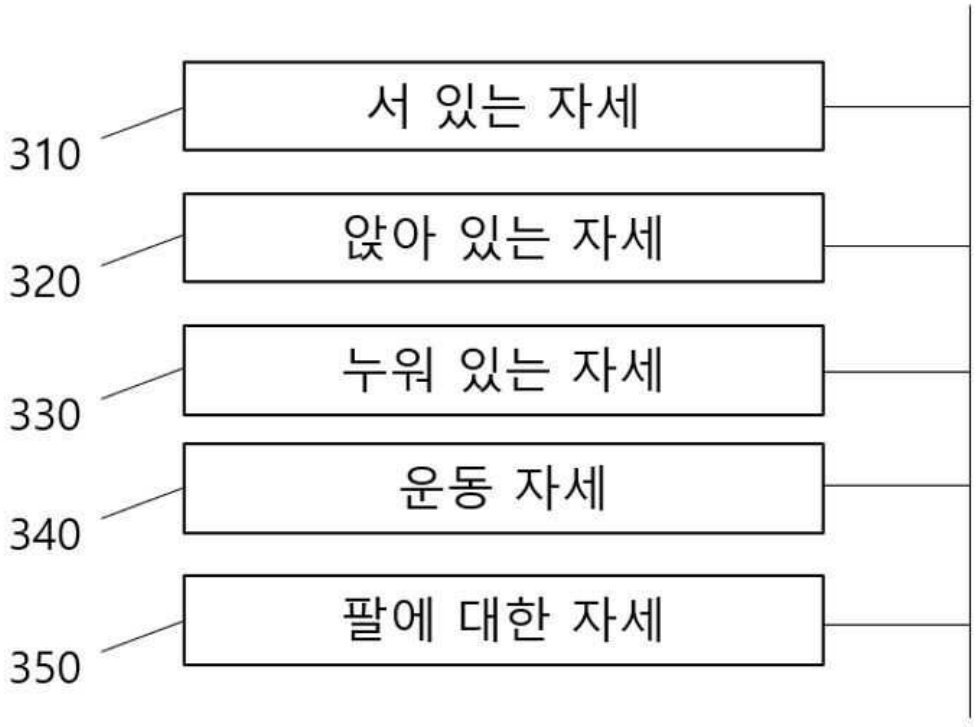
도면4


416
416
3
Conv. Layer 3 × 3 × 32
MaxPooling
208
208
32
Conv. Layer 3 × 3 × 64
MaxPooling
104
104
64
Conv. Layer 3 × 3 × 128
1 × 1 × 64
3 × 3 × 128
MaxPooling
52
52
128
Conv. Layer 3 × 3 × 256
1 × 1 × 128
3 × 3 × 256
MaxPooling
26
26
256
(a)
Conv. Layer 3 × 3 × 512
1 × 1 × 256
3 × 3 × 512
1 × 1 × 256
3 × 3 × 512(a)
MaxPooling
13
13
512
13
13
2048
Reorg / 2
Conv. Layer 3 × 3 × 1024
1 × 1 × 512
3 × 3 × 1024
1 × 1 × 512
3 × 3 × 1024
3 × 3 × 1024
13
13
3072
Conv. Layer 3 × 3 × 1024
1 × 1 × 125
13
13
125

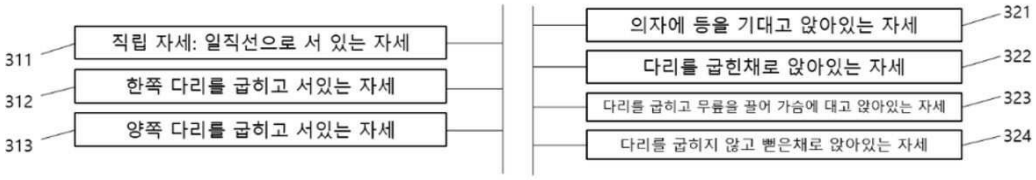
도면5


Residual Network
3x3 convolution
3x3 convolution
3x3 convolution
3x3 convolution
3x3 convolution
3x3 convolution
3x3 convolution
3x3 convolution
3x3 convolution
3x3 convolution
3x3 convolution
3x3 convolution
Softmax

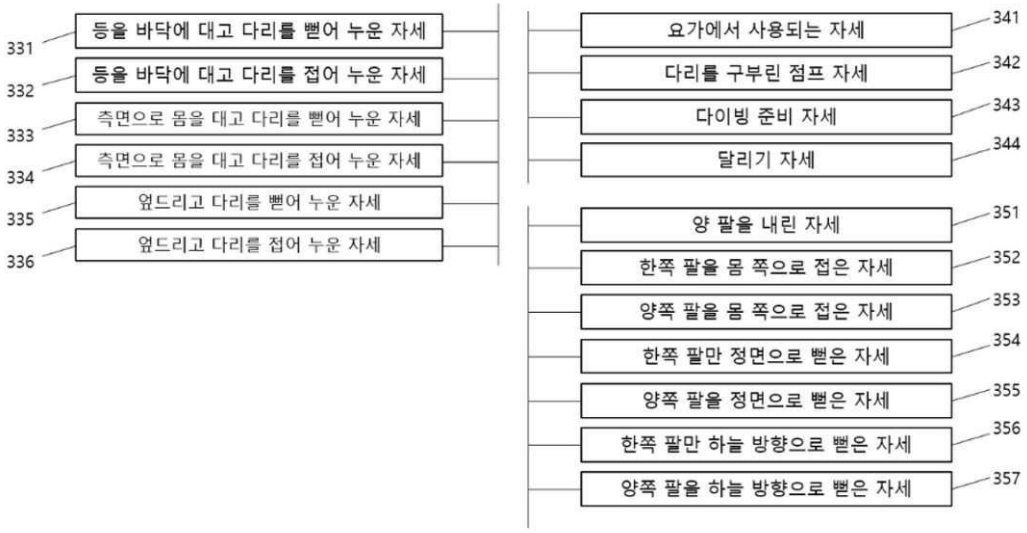
도면6


서 있는 자세
310
앉아 있는 자세
320
누워 있는 자세
330
운동 자세
340
팔에 대한 자세
350

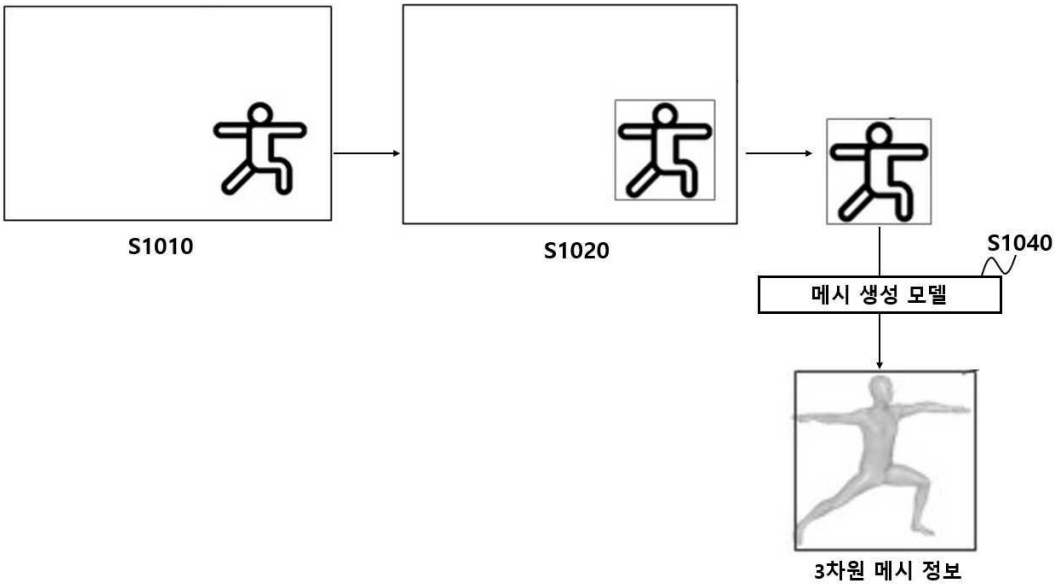
도면7


직립 자세: 일직선으로 서 있는 자세
311
한쪽 다리를 굽히고 서있는 자세
312
양쪽 다리를 굽히고 서있는 자세
313
의자에 등을 기대고 앉아있는 자세
321
다리를 굽힌채로 앉아있는 자세
322
다리를 굽히고 무릎을 팔어 가슴에 대고 앉아있는 자세
323
다리를 굽히지 않고 뻗은채로 앉아있는 자세
324

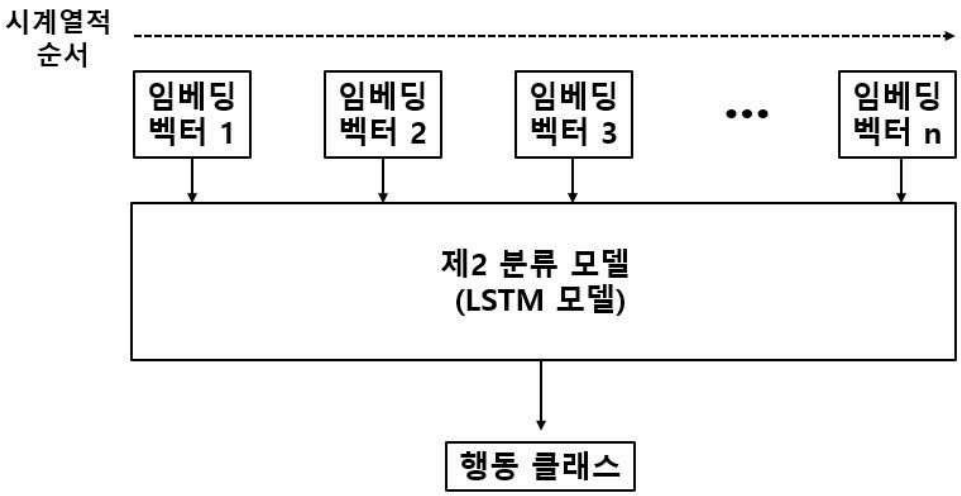
도면8


등을 바닥에 대고 다리를 뻗어 누운 자세
331
등을 바닥에 대고 다리를 접어 누운 자세
332
측면으로 몸을 대고 다리를 뻗어 누운 자세
333
측면으로 몸을 대고 다리를 접어 누운 자세
334
엎드리고 다리를 뻗어 누운 자세
335
엎드리고 다리를 접어 누운 자세
336
요가에서 사용되는 자세
341
다리를 구부린 점프 자세
342
다이빙 준비 자세
343
달리기 자세
344
양 팔을 내린 자세
351
한쪽 팔을 몸 쪽으로 접은 자세
352
양쪽 팔을 몸 쪽으로 접은 자세
353
한쪽 팔만 정면으로 뻗은 자세
354
양쪽 팔을 정면으로 뻗은 자세
355
한쪽 팔만 하늘 방향으로 뻗은 자세
356
양쪽 팔을 하늘 방향으로 뻗은 자세
357

도면9


S1010
S1020
S1040
메시 생성 모델
3차원 메시 정보

도면10


시계열적
순서
임베딩
벡터 1
임베딩
벡터 2
임베딩
벡터 3
임베딩
벡터 n
제2 분류 모델
(LSTM 모델)
행동 클래스