

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0000280

(43) 공개일자 2022년01월03일

(51) 국제특허분류(Int. Cl.)

G06T 7/246 (2017.01) G06T 7/136 (2017.01)

(52) CPC특허분류

G06T 7/246 (2017.01)

G06T 7/136 (2017.01)

(21) 출원번호 10-2020-0078039

(22) 출원일자 2020년06월25일

심사청구일자 없음

(71) 출원인

서강대학교산학협력단

서울특별시 마포구 백범로 35 (신수동, 서강대학교)

(72) 발명자

강석주

서울특별시 마포구 백범로 35(신수동)

박재서

서울특별시 마포구 백범로 35, 최양업관 419호(신수동)

허준호

서울특별시 마포구 백범로 35, 최양업관 419호(신수동)

(74) 대리인

유미특허법인

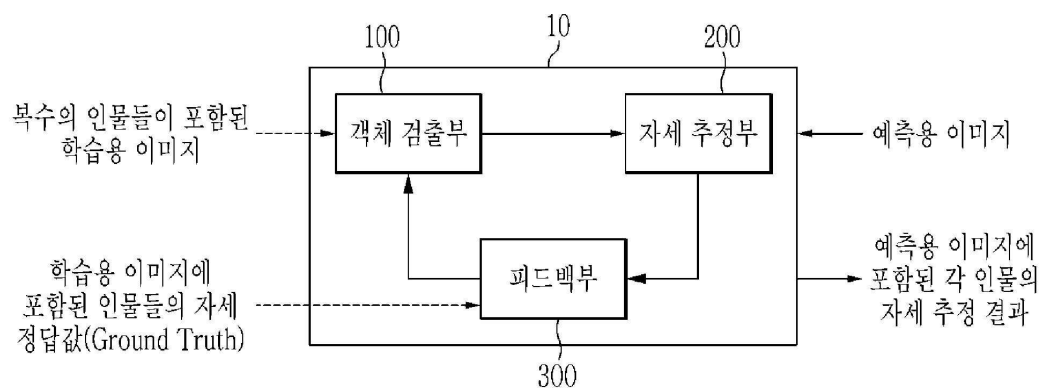
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 다중 인물의 자세 추정 방법 및 장치

(57) 요약

적어도 하나의 프로세서에 의해 동작하는 컴퓨팅 장치가 동작하는 방법으로서, 복수의 인물들이 포함된 이미지를 입력받는 단계, 학습된 인물 검출 모델을 이용하여 상기 이미지로부터 각 인물을 검출하는 단계, 학습된 자세 추정 모델을 이용하여 상기 각 인물의 신체적 특징점을 추출하고, 추출한 신체적 특징점을 이용하여 상기 각 인물의 자세를 추정하는 단계, 그리고 상기 자세 추정 모델의 정확도를 상기 인물 검출 모델로 피드백하여 상기 인물 검출 모델을 재학습시키는 단계를 포함한다.

대표도 - 도1



(52) CPC특허분류

G06T 2207/20081 (2013.01)

G06T 2207/30196 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	20PQWO-B153369-02
과제번호	202055002
부처명	행정안전부
과제관리(전문)기관명	국토교통과학기술진흥원
연구사업명	스마트 도로조명 플랫폼 개발 및 실증연구
연구과제명	스마트 도로조명 활용 도시재난안전관리 연계 기술 개발
기 여 율	1/2
과제수행기관명	서강대학교 산학협력단
연구기간	2020.01.01~2020.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711116161
과제번호	201939032
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	대학ICT연구센터지원사업
연구과제명	인공지능 서비스 실현을 위한 지능형 반도체 설계 핵심기술 개발
기 여 율	1/2
과제수행기관명	서강대학교 산학협력단
연구기간	2020.01.01~2021.12.31

명세서

청구범위

청구항 1

적어도 하나의 프로세서에 의해 동작하는 컴퓨팅 장치가 동작하는 방법으로서,
복수의 인물들이 포함된 이미지를 입력받는 단계,
학습된 인물 검출 모델을 이용하여 상기 이미지로부터 각 인물을 검출하는 단계,
학습된 자세 추정 모델을 이용하여 상기 각 인물의 신체적 특징점을 추출하고, 추출한 신체적 특징점을 이용하여 상기 각 인물의 자세를 추정하는 단계, 그리고
상기 자세 추정 모델의 정확도를 상기 인물 검출 모델로 피드백하여 상기 인물 검출 모델을 재학습시키는 단계를 포함하는 동작 방법.

청구항 2

제1항에서,
상기 검출하는 단계는,
상기 각 인물의 신체적 특징점을 포함하는 경계 상자를 생성하고, 상기 이미지에 포함된 인물의 수가 클수록 각 경계 상자의 크기는 증가하고, 각 경계 상자의 크기 증가율은 감소하는, 동작 방법.

청구항 3

제1항에서,
상기 인물 검출 모델의 손실 함수는,
상기 자세 추정 모델의 정확도가 낮을수록 상기 인물 검출 모델의 손실을 증가시키는, 동작 방법.

청구항 4

제3항에서,
상기 자세 추정 모델의 정확도는
학습용 이미지로부터 상기 학습된 자세 추정 모델이 예측한 상기 각 인물의 신체적 특징점들과, 자세 정답값(Ground Truth)의 신체적 특징점들의 거리 차이를 기초로 계산되는, 동작 방법.

청구항 5

컴퓨팅 장치로서,
메모리, 그리고
상기 메모리에 로드된 프로그램의 명령들(instructions)을 실행하는 적어도 하나의 프로세서를 포함하고,
상기 프로그램은
학습용 이미지와 상기 학습용 이미지에 포함된 복수의 인물들의 자세 정답값(Ground Truth)을 입력받는 단계,
학습된 인물 검출 모델을 이용하여 상기 학습용 이미지로부터 상기 복수의 인물들을 검출하는 단계,
각 인물의 신체적 특징점을 추출하고, 추출한 신체적 특징점을 이용하여 상기 각 인물의 자세를 추정하는 단계, 그리고
자세 추정값과 상기 자세 정답값을 비교하여 오차를 계산하고, 상기 오차를 상기 인물 검출 모델로 피드백하여 상기 인물 검출 모델을 재학습시키는 단계,

를 실행하도록 기술된 명령들을 포함하는, 컴퓨팅 장치.

청구항 6

제5항에서,

상기 재학습시키는 단계는,

상기 자세 추정값의 키포인트들과 상기 자세 정답값의 키포인트들의 거리 차이를 기초로 상기 오차를 계산하는, 컴퓨팅 장치.

청구항 7

제6항에서,

상기 재학습시키는 단계는,

상기 키포인트들이 위치한 신체 부위에 따라 거리 차이에 서로 다른 가중치가 적용되는, 컴퓨팅 장치.

청구항 8

제5항에서,

상기 재학습시키는 단계는,

상기 오차가 클수록 상기 인물 검출 모델에 포함된 손실 함수의 손실을 증가시키는, 컴퓨팅 장치.

청구항 9

제5항에서,

상기 검출하는 단계는,

상기 인물의 신체적 특징점을 포함하는 경계 상자를 생성하고,

상기 경계 상자의 크기는 상기 이미지에 포함된 인물의 수에 따라 결정되는, 컴퓨팅 장치.

청구항 10

제9항에서,

상기 검출하는 단계는,

상기 이미지에 포함된 인물의 수가 클수록 상기 경계 상자의 크기는 증가하고, 상기 경계 상자의 크기 증가율은 감소하는, 컴퓨팅 장치.

발명의 설명

기술 분야

[0001] 본 발명은 다중 인물의 자세를 추정하는 기술에 관한 것이다.

배경 기술

[0002] 복수의 인물들의 자세를 추정하는 기술(Multi-Person Pose Estimation, 이하 'MPPE' 라고 호칭함)의 핵심은 이미지에 포함된 모든 인물에 대한 키포인트(Keypoint)를 추정하는 것이다. 키포인트란 신체의 특징점을 의미한다.

[0003] MPPE는 인간 행동 인식 및 인간-컴퓨터 상호 작용과 같은 많은 시각적 응용 프로그램에서 활용된다. 한 예로서 MPPE가 차량 및 폐쇄 회로 텔레비전(CCTV)에 사용되는 경우, 속도와 정확성을 모두 고려해야 한다. 철저한 검색 기능을 갖춘 느린 객체 검출기에서 이미지 그리드의 모든 객체를 동시에 찾는 빠른 모델로 변경하려는 시도가 있었으나, 정확도가 낮아지는 단점이 있다.

[0004] 최근 딥러닝 기반 알고리즘을 이용한 연구가 객체 인식, 인간 자세 추정 및 음성 인식 분야 등 다양한 분야에서

뛰어난 성능을 보여주고 있다. 이에 따라 MPPE의 성능 역시 합성곱 신경망(Convolutional Neural Network, 이하 'CNN'이라 호칭함)에 의해 크게 개선되었다. CNN을 사용한 MPPE의 구조는 하향식 접근법(Top-down approach)과 상향식 접근법(Bottom-up approach)의 두 가지 유형으로 나뉜다.

- [0005] 상향식 접근 방식은 인물의 키포인트를 먼저 감지한 다음, 감지한 키포인트를 연결하여 인체를 형성한다. 상향식 접근 방식은 입력 이미지에 포함된 인물의 수가 증가하더라도 빠른 속도를 유지할 수 있으나, 정확도가 낮고 키포인트들을 연결하는 과정이 복잡하기 때문에 성능을 개선하기가 어렵다.
- [0006] 하향식 접근 방식은 2단계 파이프 라인으로 자세를 추정한다. 객체 검출기(Object Detector)를 통해 이미지에 포함된 모든 인물을 감지한다. 이때 인물을 포함할 가능성이 있는 상자의 위치와 규모를 예측한다. 이후 감지된 인물 각각을 자세 추정기(Single Person Pose Estimation, 이하 'SPPE'라고 호칭함)에 개별적으로 입력한다.
- [0007] SPPE는 검출된 상자 각각에서 키포인트를 예측한다. SPPE의 결과는 객체 검출기를 통해 입력되는 이미지의 품질과 해상도에 의존하기 때문에, 객체 검출기의 오류는 전체 시스템의 성능을 크게 손상시킬 수 있다.
- [0008] 구체적으로, 하향식 접근법에서는 객체 검출기의 부적절한 결과가 SPPE로 곧바로 전파된다. 객체 검출기의 성능 평가에 사용되는 일반적인 방식은 경계 상자의 가장자리에 위치한 이미지의 손실을 가져올 확률이 크다. 이는 객체 검출기에서는 적절한 결과이지만, SPPE에서는 더 큰 오류가 발생할 수 있다.
- [0009] 즉, 기존 MPPE 관련 기술은 단순히 SPPE를 객체 탐지기에 연결하고, 객체 검출기와 SPPE를 독립적으로 훈련하고 평가한다. 독립적으로 훈련된 객체 검출기는 SPPE에 더 큰 오류를 제공할 수 있다. 결과적으로, 객체 검출기와 SPPE의 결과는 긴밀하게 연관되어 있으나 객체 탐지기를 SPPE의 결과에 따라 학습시키기 위한 연구가 없었다.

발명의 내용

해결하려는 과제

- [0010] 해결하고자 하는 과제는 복수의 인물들의 자세를 추정하기 위해 사용되는 객체 검출기를 학습시키기 위해, 개별 인물의 자세를 추정하는 자세 추정기(Single Person Pose Estimation)의 결과를 피드백으로 입력하는 방법을 제공하는 것이다.
- [0011] 또한, 해결하고자 하는 과제는 이미지에 포함된 인물의 수에 따라, 인물의 키포인트를 포함하는 경계 상자의 범위를 변경시키는 방법을 제공하는 것이다.

과제의 해결 수단

- [0012] 한 실시예에 따라 적어도 하나의 프로세서에 의해 동작하는 컴퓨팅 장치가 동작하는 방법으로서, 복수의 인물들이 포함된 이미지를 입력받는 단계, 학습된 인물 검출 모델을 이용하여 상기 이미지로부터 각 인물을 검출하는 단계, 학습된 자세 추정 모델을 이용하여 상기 각 인물의 신체적 특징점을 추출하고, 추출한 신체적 특징점을 이용하여 상기 각 인물의 자세를 추정하는 단계, 그리고 상기 자세 추정 모델의 정확도를 상기 인물 검출 모델로 피드백하여 상기 인물 검출 모델을 재학습시키는 단계를 포함한다.
- [0013] 상기 검출하는 단계는, 상기 각 인물의 신체적 특징점을 포함하는 경계 상자를 생성하고, 상기 이미지에 포함된 인물의 수가 클수록 각 경계 상자의 크기는 증가하고, 각 경계 상자의 크기 증가율은 감소할 수 있다.
- [0014] 상기 인물 검출 모델의 손실 함수는, 상기 자세 추정 모델의 정확도가 낮을수록 상기 인물 검출 모델의 손실을 증가시킬 수 있다.
- [0015] 상기 자세 추정 모델의 정확도는 학습용 이미지로부터 상기 학습된 자세 추정 모델이 예측한 상기 각 인물의 신체적 특징점들과, 자세 정답값(Ground Truth)의 신체적 특징점들의 거리 차이를 기초로 계산될 수 있다.
- [0016] 한 실시예에 따른 컴퓨팅 장치로서, 메모리, 그리고 상기 메모리에 로드된 프로그램의 명령들(instructions)을 실행하는 적어도 하나의 프로세서를 포함한다. 상기 프로그램은 학습용 이미지와 상기 학습용 이미지에 포함된 복수의 인물들의 자세 정답값(Ground Truth)을 입력받는 단계, 학습된 인물 검출 모델을 이용하여 상기 학습용 이미지로부터 상기 복수의 인물들을 검출하는 단계, 각 인물의 신체적 특징점을 추출하고, 추출한 신체적 특징점을 이용하여 상기 각 인물의 자세를 추정하는 단계, 그리고 자세 추정값과 상기 자세 정답값을 비교하여 오차를 계산하고, 상기 오차를 상기 인물 검출 모델로 피드백하여 상기 인물 검출 모델을 재학습시키는 단계를 실행하도록 기술된 명령들을 포함한다.

- [0017] 상기 재학습시키는 단계는, 상기 자세 추정값의 키포인트들과 상기 자세 정답값의 키포인트들의 거리 차이를 기초로 상기 오차를 계산할 수 있다.
- [0018] 상기 재학습시키는 단계는, 상기 키포인트들이 위치한 신체 부위에 따라 거리 차이에 서로 다른 가중치가 적용될 수 있다.
- [0019] 상기 재학습시키는 단계는, 상기 오차가 클수록 상기 인물 검출 모델에 포함된 손실 함수의 손실을 증가시킬 수 있다.
- [0020] 상기 검출하는 단계는, 상기 인물의 신체적 특징점을 포함하는 경계 상자를 생성하고, 상기 경계 상자의 크기는 상기 이미지에 포함된 인물의 수에 따라 결정될 수 있다.
- [0021] 상기 검출하는 단계는, 상기 이미지에 포함된 인물의 수가 클수록 상기 경계 상자의 크기는 증가하고, 상기 경계 상자의 크기 증가율은 감소할 수 있다.

발명의 효과

- [0022] 본 발명에 따르면, 2단계로 이루어지는 자세 추정 과정에서 개별 인물의 자세 추정 결과를 이용하여 객체 검출기를 학습시키므로, 독립적으로 훈련된 객체 검출기에 비해 자세 추정의 최종 결과의 정확도를 높일 수 있다.
- [0023] 또한 본 발명에 따르면, 객체를 검출하기 위한 경계 상자의 범위가 고정된 값이 아니고 이미지에 따라 변하므로, 각 인물의 키포인트를 누락하지 않고 추출할 수 있다.

도면의 간단한 설명

- [0024] 도 1은 한 실시예에 따른 자세 추정 장치의 구성도이다.
- 도 2는 한 실시예에 따른 자세 추정 장치의 동작 방법의 예시도이다.
- 도 3은 한 실시예에 따른 객체 검출부의 동작 방법의 예시도이다.
- 도 4는 한 실시예에 따른 피드백부의 동작 방법의 예시도이다.
- 도 5는 한 실시예에 따른 자세 추정 장치의 동작 방법의 흐름도이다.
- 도 6과 도 7은 한 실시예에 따른 자세 추정 장치의 성능을 나타낸 그래프이다.
- 도 8은 한 실시예에 따른 컴퓨팅 장치의 하드웨어 구성도이다.

발명을 실시하기 위한 구체적인 내용

- [0025] 아래에서는 첨부한 도면을 참고로 하여 본 발명의 실시예에 대하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.
- [0026] 명세서 전체에서, 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다. 또한, 명세서에 기재된 "...부", "...기", "모듈" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.
- [0027] 본 명세서에서, 설명의 편의상 인물을 객체로 호칭한다.
- [0028] 도 1은 한 실시예에 따른 자세 추정 장치의 구성도이고, 도 2는 한 실시예에 따른 자세 추정 장치의 동작 방법의 예시도이고, 도 3은 한 실시예에 따른 객체 검출부의 동작 방법의 예시도이고, 도 4는 한 실시예에 따른 피드백부의 동작 방법의 예시도이다.
- [0029] 도 1을 참고하면, 자세 추정 장치(10)는 객체 검출부(100), 자세 추정부(200) 그리고 피드백부(300)를 포함한다.
- [0030] 객체 검출부(100), 자세 추정부(200) 그리고 피드백부(300)는 하나의 컴퓨팅 장치에 구현되거나, 별도의 컴퓨팅

장치에 분산 구현될 수 있다. 별도의 컴퓨팅 장치에 분산 구현된 경우, 객체 검출부(100), 자세 추정부(200) 그리고 피드백부(300)는 통신 인터페이스를 통해 서로 통신할 수 있다. 컴퓨팅 장치는 본 발명을 수행하도록 작성된 소프트웨어 프로그램을 실행할 수 있는 장치이면 충분하고, 예를 들면, 서버, 랩탑 컴퓨터 등일 수 있다.

[0031] 객체 검출부(100), 자세 추정부(200) 그리고 피드백부(300) 각각은 하나의 인공지능 모델일 수 있고, 복수의 인공지능 모델로 구현될 수도 있다. 자세 추정 장치(10)는 하나의 인공지능 모델일 수 있고, 복수의 인공지능 모델로 구현될 수도 있다. 이에 따라, 상술한 구성들에 대응하는 하나 또는 복수의 인공지능 모델은 하나 또는 복수의 컴퓨팅 장치에 의해 구현될 수 있다.

[0032] 도 2를 참고하면, 자세 추정 장치(10)는 도 2의 (a)와 같은 복수의 인물들이 포함된 이미지를 입력받는다. 객체 검출부(100)는 도 2의 (a)로부터 2명의 인물을 검출하고 그 결과는 도 2의 (b)와 같을 수 있다. 객체 검출부(100)가 도 2의 (b)를 자세 추정부(200)로 전달하면, 자세 추정부(200)는 각 경계 상자에 포함된 인물들의 신체적 특징점을 검출하여 자세를 추정한다. 그 결과는 도 2의 (c)와 같을 수 있다.

[0033] 객체 검출부(100)는 복수의 인물들이 포함된 이미지를 입력받고, 이미지에 포함된 인물, 즉 객체들을 검출하고 객체들의 위치를 추출하여 자세 추정부(200)에 전달한다. 객체 검출부(100)는 복수의 객체 각각의 위치를 파악할 수 있다. 객체 검출부(100)는 이미지에서 검출한 객체들을 사각형 상자로 표시할 수 있으며, 이를 윈도우(Window) 또는 경계 상자(Bounding Box)라고 호칭할 수 있다. 한편 이미지에서 특정 객체를 검출하는 기술은 이미 공지된 기술이므로 자세한 설명은 생략한다.

[0034] 객체 검출부(100)는 머신러닝과 분류기의 결합 또는 딥러닝 모델로 구현될 수 있으며, 한 예로서 컨벌루션 뉴럴 네트워크(Convolutional Neural Network, CNN)으로 구현될 수 있다. 다른 예로서, YOLO(You Only Look Once) 알고리즘으로 구현될 수 있다.

[0035] YOLO란, 입력되는 이미지를 동일한 크기의 그리드로 나누고, 각 그리드에서 미리 정의된 형태(Predefined Shape)로 지정된 경계 박스의 개수를 예측한다. 각 경계 박스의 신뢰도를 계산한 후 높은 객체 신뢰도를 가진 위치를 선택하는 단일 단계 방식의 객체 탐지 알고리즘을 의미한다.

[0036] 객체 검출부(100)가 딥러닝 모델로 구현된 경우, 객체 검출부(100)의 학습 과정에 사용되는 손실 함수는 자세 추정 장치(10)의 최종 출력값에 의해 수정될 수 있다.

[0037] 도 3을 참고하면, 경계 상자의 크기는 조절될 수 있다. 구체적으로, 객체 검출부(100)는 도 3의 (a)와 같이 이미지에서 키포인트 정보가 포함된 영역을 누락하지 않도록, 도 3의 (b)와 같이 경계 상자를 확장할 수 있다. 한 예로서, 경계 상자의 크기는 이미지에 포함된 객체의 수에 따라 조절될 수 있다. 도 2를 참고하면, 예측1과 예측2의 경계 상자는 모두 다르지만, IoU(Intersection over Union) 값은 모두 동일하다.

[0038] 이미지에 포함된 객체의 수가 클수록, 하나의 경계 상자 안에 다른 객체가 포함될 가능성이 높아진다. 따라서 이미지에 많은 인물이 포함된 것일수록, 경계 상자의 크기가 확장되는 비율이 감소해야 한다. 경계 상자 크기의 확장 비율을 α 라고 하면, α 는 수학식 1에 의해 계산될 수 있다.

수학식 1

$$\alpha = \log(n(p))\gamma + \mu, \quad p \in P$$

[0039]

[0040] 수학식 1에서, α 는 경계 상자가 얼마나 확장될 것인지를 의미하며, p 의 개수에 따라 감소하는 단조 함수이다. 즉 α 는 검출된 객체의 수를 기반으로 경계 상자를 확장할 범위를 특징하는 확장 비율을 의미한다. P 는 객체 검출부(100)에서 출력되는 모든 경계 상자들을 의미하고, p 는 하나의 이미지에서 출력되는 경계 상자들을 의미한다. γ 와 μ 는 확장 변수 α 가 얼마나 급격하게 반영되는지를 조절하는 하이퍼 파라미터이며, 하이퍼 파라미터들의 값은 실험 결과에 따라 결정될 수 있다.

[0041] α 를 반영하여 최종적으로 수정된 경계 상자의 크기는 수학식 2와 같다.

수학식 2

$$1 + \frac{1}{\alpha}$$

[0042]

[0043]

수학식 2에 의해 계산된 경계 상자의 크기는 도 7의 그래프와 같다.

[0044]

자세 추정부(200)는 객체 검출부(100)로부터 전달받은 복수의 경계 상자들에 포함된 각 인물들의 신체적 특징점인 키포인트들을 추출하고, 키포인트를 바탕으로 각 인물들의 자세를 추정한다. 또한 자세 추정부(200)는 추출한 키포인트들을 피드백부(300)로 전달한다.

[0045]

인물의 자세를 추정하기 위해 사용되는 알고리즘은 어느 하나로 제한되지 않는다. 예를 들어 공지된 Fast Human Pose Estimation 알고리즘을 사용할 수 있다.

[0046]

피드백부(300)는 자세 추정부(200)로부터 각 인물들의 키포인트를 전달받고, 각 인물들 자세의 정답값(Ground Truth)과 비교하여 자세 추정부(200)가 예측한 키포인트의 정확도 또는 성능을 평가하고, 정확도에 따라 객체 검출부(100)의 손실 함수를 수정할 수 있다.

[0047]

일반적으로 손실 함수는, 예측된 출력값과 정답값의 오차 또는 괴리도를 정의하는 함수로서, 오차제곱합(Sum of Squares for Error, SSE), 교차 엔트로피 오차(Cross Entropy Error, CEE)가 주로 사용된다. 최적화 과정을 통해 손실이 최소화 되는 지점을 찾는 것이 딥러닝의 목표이므로, 딥러닝 모델의 성능을 높이기 위해 손실 함수를 변경할 수 있는 것이다.

[0048]

도 4를 참고하면, 자세 추정부(200)의 출력인 도 4의 (a)에 포함된 키포인트와 정답값인 도 4의 (b)에 포함된 키포인트를 비교할 수 있다.

[0049]

성능을 평가하는 방법으로서, 자세 추정부(200)의 결과인 예측값과 정답값 간의 거리 차이를 이용할 수 있다. 한 예로서, 객체 키포인트 유사성(Object Keypoint Similarity, OKS)이라는 변수일 수 있다. 객체 키포인트 유사성은 키포인트 유형에 따라 달라지는 값일 수 있다.

[0050]

객체 키포인트 유사성은 수학식 3에 의해 계산될 수 있으며, 공지된 기술인바 자세한 설명은 생략한다.

수학식 3

$$OKS = \exp\left(-\frac{d_i^2}{2s^2k_i^2}\right)$$

[0051]

[0052]

수학식 3에서, d_i 예측된 키포인트와 정답 키포인트 간 유클리디안 거리를 의미하고, s 는 경계 상자의 영역을 전체 이미지 영역으로 나눈 스케일 값, k 는 키포인트 상수를 의미한다.

[0053]

한편 피드백부(300)는, 신체 부위에 따라 객체 키포인트 유사성을 다르게 계산할 수 있다. 예를 들어 인물의 눈에 해당하는 영역은 인물의 어깨 또는 엉덩이 영역보다 중요한 부분이므로, 눈에 해당하는 키포인트가 정확히 예측되었는지는 다른 부위에 비해 엄격히 평가될 수 있다.

[0054]

피드백부(300)는 자세 추정부(200)의 정확도를 평가하고, 정확도를 기반으로 객체 검출부(100)의 손실 함수를 수정한다. 수정된 손실 함수는 수학식 4와 같다.

수학식 4

$$L_f = \beta((\ln(x))^2), x \in D$$

[0055]

- [0056] 수학식 4에서, L_f 는 객체 검출부(100)의 손실 함수를 의미하고, β 는 손실 함수의 페널티의 강도를 조절하기 위한 매개 변수이며, β 값은 관리자에 의해 임의로 설정될 수 있다. D 는 각 이미지 별 예측 정밀도를 의미한다.
- [0057] 즉, 피드백부(300)는 자세 추정부(200)의 결과를 이용하여 겹치는 객체나 군중 객체가 입력되더라도 객체 검출부(100)에서 세밀하고 강건하게 결과를 출력할 수 있도록 민감도를 수정할 수 있다. 결과적으로 자세 추정부(200)의 예측 오차가 클수록, 객체 검출부(100)의 손실 값을 증가시킨다.
- [0058] 각 이미지의 손실 함수 스코어는 수정된 손실 함수에 의해 계산된다. 수정된 손실 함수는 단독으로 사용되지 않고, 기존의 평균 제곱 오차(Mean Squared Error, MSE)를 조정하는 형태로 사용된다. 이로서 객체 검출부(100)와 자세 추정부(200)의 병합은 직접적으로 일어나지 않는다.
- [0059] 도 5는 한 실시예에 따른 자세 추정 장치의 동작 방법의 흐름도이다.
- [0060] 도 5를 참고하면, 자세 추정 장치(10)는 복수의 인물들이 포함된 학습용 이미지와 학습용 이미지에 포함된 인물들의 자세 정답값(Ground Truth)을 입력받는다(S110).
- [0061] 객체 검출부(100)는 학습용 이미지에 포함된 복수의 객체들을 경계 상자 형태로 탐지한다(S120). 이때, 객체 검출부(100)는 이미지에 포함된 객체의 수에 따라 경계 상자의 크기를 조절할 수 있으며, 이는 수학식 1을 통해 설명한 바와 같다.
- [0062] 자세 추정부(200)는 각 경계 상자에 포함된 객체의 키포인트를 예측한다(S130).
- [0063] 피드백부(300)는 자세 추정부(200)가 예측한 키포인트와 정답값과의 오차를 계산한다(S140). 정답값은 학습용 이미지와 같이 입력될 수 있으며, 거리 기반의 오차는 수학식 2를 통해 계산될 수 있다.
- [0064] 피드백부(300)는 S140 단계에서 계산된 오차를 기반으로 객체 검출부(100)의 손실 함수를 수정한다(S150). 구체적으로, 자세 추정부(200)의 예측 오차가 클수록 객체 검출부(100)의 손실 값이 증가하도록 객체 검출부(100)의 손실 함수를 수정한다.
- [0065] 객체 검출부(100)는 피드백부(300)에 의해 수정된 손실 함수를 이용하여 학습을 진행한다(S160). 이후 자세 추정 장치(10)는 예측용 이미지를 입력받고, 예측용 이미지에 포함된 인물들을 검출하고, 각 인물의 자세를 추정하여 출력할 수 있다.
- [0066] 도 6과 도 7은 한 실시예에 따른 자세 추정 장치의 성능을 나타낸 그래프이다.
- [0067] 도 6과 도 7을 참고하면, 자세 추정 장치(10)에 포함된 매개 변수 또는 하이퍼 파라미터들을 변경하면서, 학습과 평가를 수행한 결과를 나타낸 것이다.
- [0068] 도 6은 객체 검출부(100)의 손실 함수에서, 수학식 3에 포함된 변수 β 값을 조정하며 손실 값 즉, 페널티를 측정한 것이다. 피드백부(300)에 의해 계산된 객체 키포인트 유사성 값이 0.5 미만일수록 손실 값이 빠르게 증가한다.
- [0069] 또한 변수 β 에 따른 자세 추정 장치(10)의 정확도는 아래의 표 1과 같다. 각 β 값에 대해 복수 회 평가 후 평균(Average Precision)을 측정한 것이다.

표 1

변수 β 의 값	정확도(AP)
10	72.8
5	73.0
2	73.1

- [0070]
- [0071] 도 7은 객체 검출부(100)가 수학식 1에 의해 경계 상자를 확장할 때, 하이퍼 파라미터 γ , μ 에 따른 경계 상자의 크기를 나타낸 것이다. 도 7을 참고하면, 이미지에 포함된 인물의 수가 증가할수록 경계 상자의 크기 역시 증가하나, 경계 상자 크기의 확장률은 감소함을 알 수 있다.
- [0072] 또한 γ , μ 값을 변화시키며 자세 추정 장치(10)의 정확도를 평가한 결과는 표 2와 같다.

표 2

변수 γ 의 값	변수 μ 의 값	정확도(AP)
4	6	73.6
2	4	73.5
2	3	73.5

[0073]

[0074]

[0075]

[0076]

[0077]

[0078]

[0079]

[0080]

[0081]

도 8은 한 실시예에 따른 컴퓨팅 장치의 하드웨어 구성도이다.

도 8을 참고하면, 자세 추정 장치(10)는 적어도 하나의 프로세서에 의해 동작하는 컴퓨팅 장치(400)에서, 본 발명의 동작을 실행하도록 기술된 명령들(instructions)이 포함된 프로그램을 실행한다.

컴퓨팅 장치(400)의 하드웨어는 적어도 하나의 프로세서(410), 메모리(420), 스토리지(430), 통신 인터페이스(440)를 포함할 수 있고, 버스를 통해 연결될 수 있다. 이외에도 입력 장치 및 출력 장치 등의 하드웨어가 포함될 수 있다. 컴퓨팅 장치(400)는 프로그램을 구동할 수 있는 운영 체제를 비롯한 각종 소프트웨어가 탑재될 수 있다.

프로세서(410)는 컴퓨팅 장치(400)의 동작을 제어하는 장치로서, 프로그램에 포함된 명령들을 처리하는 다양한 형태의 프로세서(410)일 수 있고, 예를 들면, CPU(Central Processing Unit), MPU(Micro Processor Unit), MCU(Micro Controller Unit), GPU(Graphic Processing Unit) 등 일 수 있다. 메모리(420)는 본 발명의 동작을 실행하도록 기술된 명령들이 프로세서(410)에 의해 처리되도록 해당 프로그램을 로드한다. 메모리(420)는 예를 들면, ROM(read only memory), RAM(random access memory) 등 일 수 있다. 스토리지(430)는 본 발명의 동작을 실행하는데 요구되는 각종 데이터, 프로그램 등을 저장한다. 통신 인터페이스(440)는 유/무선 통신 모듈일 수 있다.

본 발명에 따르면, 2단계로 이루어지는 자세 추정 과정에서 개별 인물의 자세 추정 결과를 이용하여 객체 검출기를 학습시키므로, 독립적으로 훈련된 객체 검출기에 비해 자세 추정의 최종 결과의 정확도를 높일 수 있다.

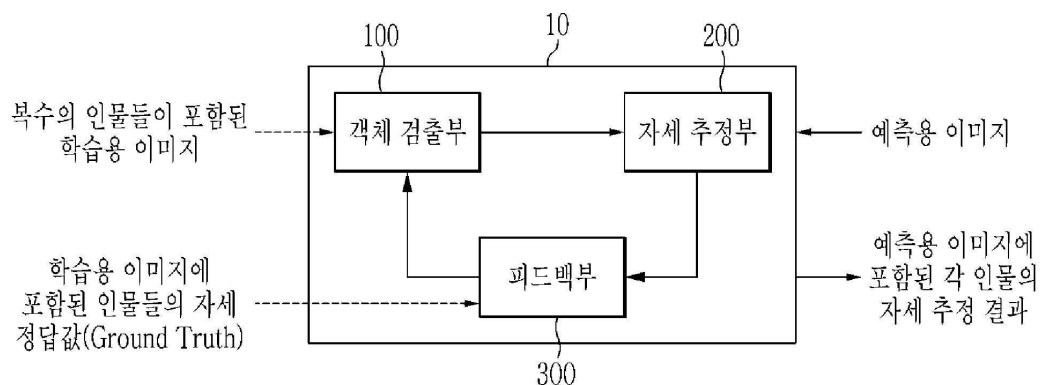
또한 본 발명에 따르면, 객체를 검출하기 위한 경계 상자의 범위가 고정된 값이 아니고 이미지에 따라 변하므로, 각 인물의 키포인트를 누락하지 않고 추출할 수 있다.

이상에서 설명한 본 발명의 실시예는 장치 및 방법을 통해서만 구현이 되는 것은 아니며, 본 발명의 실시예의 구성에 대응하는 기능을 실현하는 프로그램 또는 그 프로그램이 기록된 기록 매체를 통해 구현될 수도 있다.

이상에서 본 발명의 실시예에 대하여 상세하게 설명하였지만 본 발명의 권리범위는 이에 한정되는 것은 아니고 다음의 청구범위에서 정의하고 있는 본 발명의 기본 개념을 이용한 당업자의 여러 변형 및 개량 형태 또한 본 발명의 권리범위에 속하는 것이다.

도면

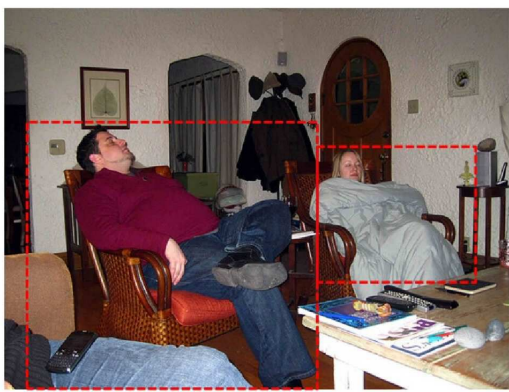
도면1



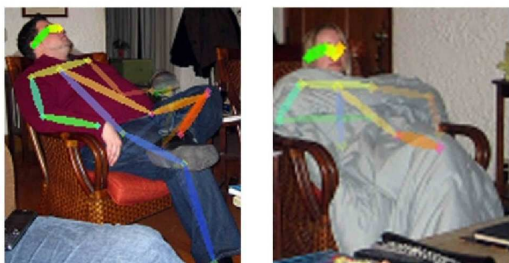
도면2



(a)

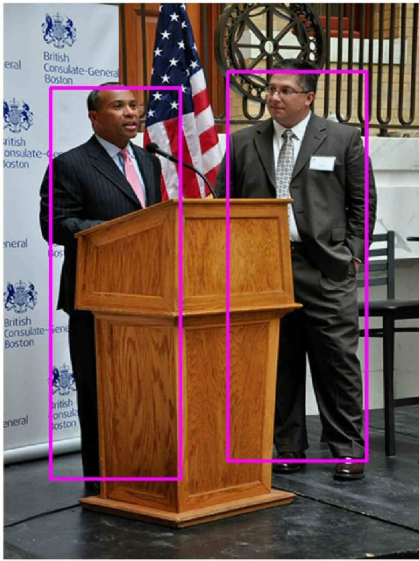


(b)



(c)

도면3



(a)



(b)

도면4

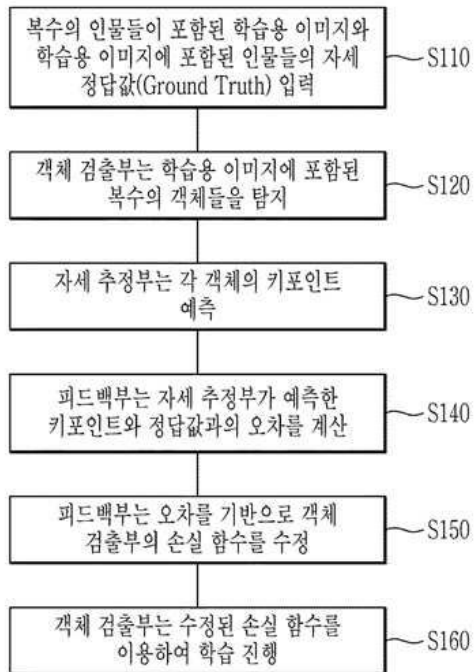


(a)

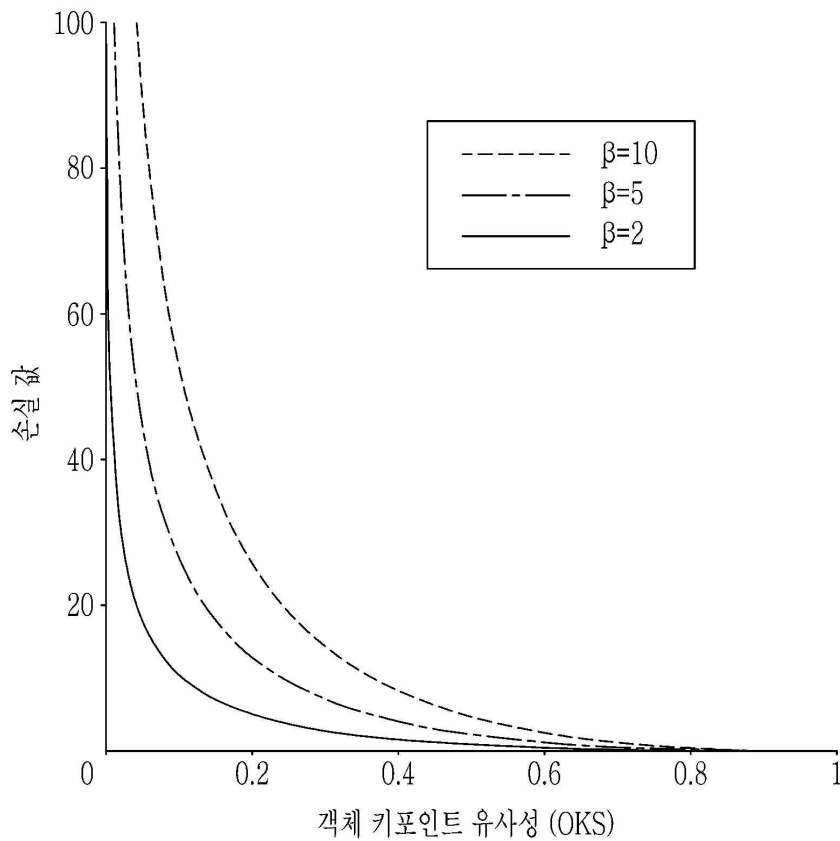


(b)

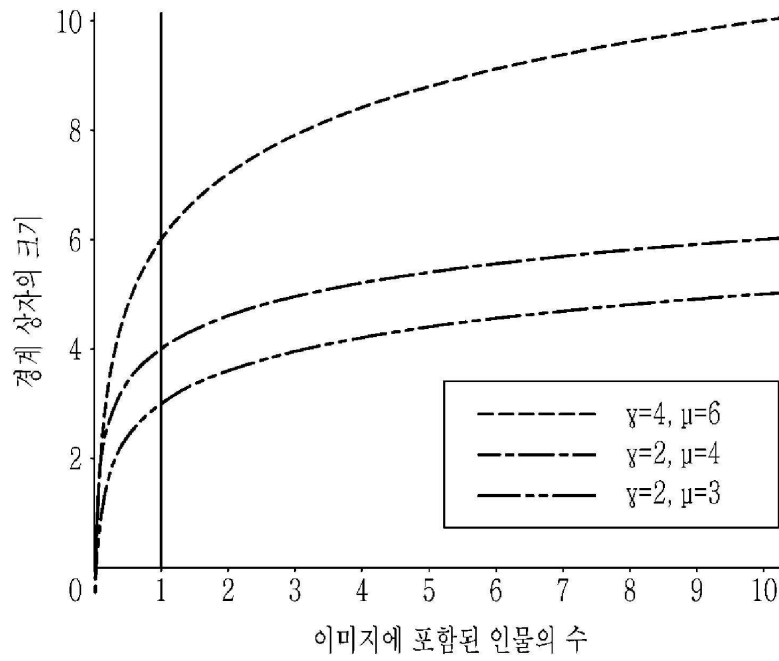
도면5



도면6



도면7



도면8

