

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)(11) 공개번호 10-2023-0057765
(43) 공개일자 2023년05월02일

(51) 국제특허분류(Int. Cl.)
G06T 7/246 (2017.01) G06N 20/00 (2019.01)
G06T 7/187 (2017.01) G06T 7/70 (2017.01)
(52) CPC특허분류
G06T 7/246 (2017.01)
G06N 20/00 (2021.08)
(21) 출원번호 10-2021-0141928
(22) 출원일자 2021년10월22일
심사청구일자 2021년10월22일

(71) 출원인
계명대학교 산학협력단
대구광역시 달서구 달구벌대로 1095, 계명대학교
산학협력관 201호(신당동)
(72) 발명자
고병철
대구광역시 수성구 범어천로 107 범어 STX KAN,
101동 1002호
김상원
대구광역시 달서구 달구벌대로 1095 계명대학교
(74) 대리인
김건우

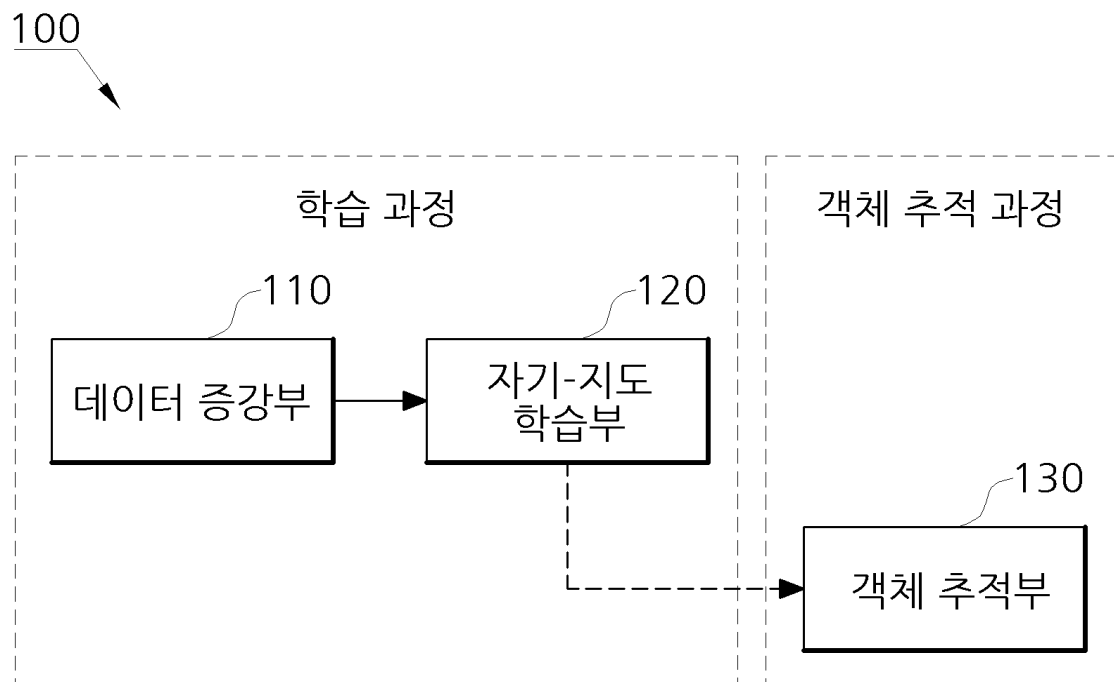
전체 청구항 수 : 총 16 항

(54) 발명의 명칭 자기-지도 학습 기반의 다중 객체 추적 장치 및 방법

(57) 요약

본 발명은 자기-지도 학습 기반의 다중 객체 추적 장치에 관한 것으로서, 보다 구체적으로는 다중 객체 추적 장치로서, 사전 학습된 포즈 정규화-생성적 적대 네트워크(pose normalized GAN(Generative Adversarial Network); PN-GAN)를 사용해, 입력 이미지로부터 포즈 변화를 생성해 복수의 포즈 이미지를 획득하며, 상기 입력 (뒷면에 계속)

대표도 - 도1



이미지와 포즈 이미지에 이미지 증강(image augmentation)을 적용해 증강 이미지를 생성하는 데이터 증강부; 상기 증강 이미지 중 양의 쌍(positive pairs)을 사용해, 다중 객체 추적(Multi-Object Tracking; MOT)을 위한 대조적 구조(contrastive structure)의 인코더 네트워크를 자기-지도 학습(Self-Supervised Learning; SSL) 방식으로 학습하는 자기-지도 학습부; 및 상기 자기-지도 학습부에서 학습된 인코더 네트워크를 사용해 추적 객체와 검출 객체 간의 유사성을 추정하고 객체 추적을 수행하는 객체 추적부를 포함하는 것을 그 구성상의 특징으로 한다.

또한, 본 발명은 자기-지도 학습 기반의 다중 객체 추적 방법에 관한 것으로서, 보다 구체적으로는 다중 객체 추적 방법으로서, (1) 사전 학습된 포즈 정규화-생성적 적대 네트워크(pose normalized GAN(Generative Adversarial Network); PN-GAN)를 사용해, 입력 이미지로부터 포즈 변화를 생성해 복수의 포즈 이미지를 획득하며, 상기 입력 이미지와 포즈 이미지에 이미지 증강(image augmentation)을 적용해 증강 이미지를 생성하는 데이터 증강 단계; (2) 상기 증강 이미지 중 양의 쌍(positive pairs)을 사용해, 다중 객체 추적(Multi-Object Tracking; MOT)을 위한 대조적 구조(contrastive structure)의 인코더 네트워크를 자기-지도 학습(Self-Supervised Learning; SSL) 방식으로 학습하는 자기-지도 학습 단계; 및 (3) 상기 단계 (2)에서 학습된 인코더 네트워크를 사용해 추적 객체와 검출 객체 간의 유사성을 추정하고 객체 추적을 수행하는 객체 추적 단계를 포함하는 것을 그 구성상의 특징으로 한다.

본 발명에서 제안하고 있는 자기-지도 학습 기반의 다중 객체 추적 장치 및 방법에 따르면, 자기-지도 학습을 온라인 다중 객체 추적에 적용한 최초의 시도이며, 객체의 포즈 변화를 고려한 학습을 위해 PN-GAN(pose normalized GAN)과 이미지 증강을 적용하고, PN-GAN을 통해 학습 데이터가 부족한 문제를 해결해 적은 수의 데이터로도 더 나은 결과를 얻을 수 있는 MOT 모델을 설계할 수 있으며, 새로운 손실 함수 ACD(Affinity Correlation Distance)를 제안하여 인코더 네트워크의 학습 시 더 많은 구별 특징 벡터를 생성할 수 있고, end-to-end 방식을 사용하여 학습할 수 있으며, 분류, 클러스터링 및 회귀에 대한 추가 미세 조정이 필요하지 않기 때문에 추가적인 전이 학습 없이 바로 적용이 가능하고, 다른 최신 방법들에 비해 적은 메모리를 차지하며 빠르고 높은 추적 성능을 제공할 수 있다.

(52) CPC특허분류

G06T 7/187 (2017.01)

G06T 7/70 (2017.01)

G06T 2207/20081 (2013.01)

G06T 2207/30232 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1345313705
과제번호	2019R111A3A01042506
부처명	교육부
과제관리(전문)기관명	한국연구재단
연구사업명	지역대학우수과학자지원사업(1년~5년)
연구과제명	역전파가 필요 없는 심층 랜덤 네트워크 모델 개발과 응용 연구
기 여 율	1/1
과제수행기관명	계명대학교
연구기간	2021.03.01 ~ 2022.02.28

명세서

청구범위

청구항 1

다중 객체 추적 장치(100)로서,

사전 학습된 포즈 정규화-생성적 적대 네트워크(pose normalized GAN(Generative Adversarial Network); PN-GAN)를 사용해, 입력 이미지로부터 포즈 변화를 생성해 복수의 포즈 이미지를 획득하며, 상기 입력 이미지와 포즈 이미지에 이미지 증강(image augmentation)을 적용해 증강 이미지를 생성하는 데이터 증강부(110);

상기 증강 이미지 중 양의 쌍(positive pairs)을 사용해, 다중 객체 추적(Multi-Object Tracking; MOT)을 위한 대조적 구조(contrastive structure)의 인코더 네트워크를 자기-지도 학습(Self-Supervised Learning; SSL) 방식으로 학습하는 자기-지도 학습부(120); 및

상기 자기-지도 학습부(120)에서 학습된 인코더 네트워크를 사용해 추적 객체와 검출 객체 간의 유사성을 추정하고 객체 추적을 수행하는 객체 추적부(130)를 포함하는 것을 특징으로 하는, 자기-지도 학습 기반의 다중 객체 추적 장치(100).

청구항 2

제1항에 있어서, 상기 자기-지도 학습부(120)는,

대조적 구조의 두 개의 인코더 네트워크를 포함하여 SSL-MOT를 구성하며, 상기 SSL-MOT는 상기 두 개의 인코더 네트워크 각각을 학습하는 제1 브랜치 및 제2 브랜치로 구성되는 것을 특징으로 하는, 자기-지도 학습 기반의 다중 객체 추적 장치(100).

청구항 3

제2항에 있어서,

상기 제1 브랜치의 인코더 네트워크와 상기 제2 브랜치의 인코더 네트워크는 가중치를 공유하는 것을 특징으로 하는, 자기-지도 학습 기반의 다중 객체 추적 장치(100).

청구항 4

제2항에 있어서, 상기 자기-지도 학습부(120)는,

상기 두 개의 증강 이미지를 상기 제1 브랜치 및 제2 브랜치에 각각 입력해 상기 SSL-MOT를 학습하되, 상기 두 개의 증강 이미지 쌍의 출력 배치 행렬에서 계산된 교차 상관(cross-correlation)을 불변 항(invariance term) 및 중복 감소 항(redundancy reduction term)과 결합해 상기 두 개의 증강 이미지 간의 유사성을 계산하는 연관성 상관 거리(Affinity Correlation Distance; ACD)를 사용하는 손실 함수를 이용해 상기 SSL-MOT를 학습을 하는 것을 특징으로 하는, 자기-지도 학습 기반의 다중 객체 추적 장치(100).

청구항 5

제4항에 있어서, 상기 연관성 상관 거리는,

상기 제1 브랜치 및 제2 브랜치에서 각각 생성된 두 개의 배치 행렬의 내적을 통해 연관성 상관 행렬(affinity correlation matrix; A)을 획득하고, 상기 연관성 상관 행렬로 상기 불변 항 및 중복 감소 항을 구성해 산출하는 것을 특징으로 하는, 자기-지도 학습 기반의 다중 객체 추적 장치(100).

청구항 6

제2항에 있어서, 상기 자기-지도 학습부(120)는,

확률적 경사 최적화기(stochastic gradient optimizer; SGD)를 적용해 학습 중 평균 손실을 역전파하되, 정지-기울기(stop-gradient)를 사용해 상기 제2 브랜치의 인코더 네트워크의 매개변수를 업데이트하지 않는 것을 특징으로 하는, 자기-지도 학습 기반의 다중 객체 추적 장치(100).

청구항 7

제1항에 있어서, 상기 객체 추적부(130)는,

복수의 추적 객체로 구성된 추적 세트와 복수의 검출 객체로 구성된 검출 세트를 각각 구성하고, 상기 추적 세트와 검출 세트의 모든 요소를 상기 자기-지도 학습부(120)에 의해 학습된 인코더 네트워크에 각각 입력해 획득한 복수의 추적 특징 벡터와 검출 특징 벡터의 쌍별 유사도를 기반으로 연관성 상관 행렬(A)을 추정하여 최종 검출-추적 객체 쌍을 설정하는 것을 특징으로 하는, 자기-지도 학습 기반의 다중 객체 추적 장치(100).

청구항 8

제7항에 있어서, 상기 객체 추적부(130)는,

데이터 연결 및 헝가리안 매칭을 통해 상기 설정한 최종 검출-추적 객체 쌍의 검출 객체를 결합해 추적 객체 상태를 업데이트하는 것을 특징으로 하는, 자기-지도 학습 기반의 다중 객체 추적 장치(100).

청구항 9

다중 객체 추적 방법으로서,

- (1) 사전 학습된 포즈 정규화-생성적 적대 네트워크(pose normalized GAN(Generative Adversarial Network); PN-GAN)를 사용해, 입력 이미지로부터 포즈 변화를 생성해 복수의 포즈 이미지를 획득하며, 상기 입력 이미지와 포즈 이미지에 이미지 증강(image augmentation)을 적용해 증강 이미지를 생성하는 데이터 증강 단계;
- (2) 상기 증강 이미지 중 양의 쌍(positive pairs)을 사용해, 다중 객체 추적(Multi-Object Tracking; MOT)을 위한 대조적 구조(contrastive structure)의 인코더 네트워크를 자기-지도 학습(Self-Supervised Learning; SSL) 방식으로 학습하는 자기-지도 학습 단계; 및
- (3) 상기 단계 (2)에서 학습된 인코더 네트워크를 사용해 추적 객체와 검출 객체 간의 유사성을 추정하고 객체 추적을 수행하는 객체 추적 단계를 포함하는 것을 특징으로 하는, 자기-지도 학습 기반의 다중 객체 추적 방법.

청구항 10

제9항에 있어서, 상기 단계 (2)에서는,

대조적 구조의 두 개의 인코더 네트워크를 포함하여 SSL-MOT를 구성하며, 상기 SSL-MOT는 상기 두 개의 인코더 네트워크 각각을 학습하는 제1 브랜치 및 제2 브랜치로 구성되는 것을 특징으로 하는, 자기-지도 학습 기반의 다중 객체 추적 방법.

청구항 11

제10항에 있어서,

상기 제1 브랜치의 인코더 네트워크와 상기 제2 브랜치의 인코더 네트워크는 가중치를 공유하는 것을 특징으로 하는, 자기-지도 학습 기반의 다중 객체 추적 방법.

청구항 12

제10항에 있어서, 상기 단계 (2)에서는,

상기 두 개의 증강 이미지를 상기 제1 브랜치 및 제2 브랜치에 각각 입력해 상기 SSL-MOT를 학습하되, 상기 두 개의 증강 이미지 쌍의 출력 배치 행렬에서 계산된 교차 상관(cross-correlation)을 불변 항(invariance term) 및 중복 감소 항(redundancy reduction term)과 결합해 상기 두 개의 증강 이미지 간의 유사성을 계산하는 연관성 상관 거리(Affinity Correlation Distance; ACD)를 사용하는 손실 함수를 이용해 상기 SSL-MOT를 학습을 하는 것을 특징으로 하는, 자기-지도 학습 기반의 다중 객체 추적 방법.

청구항 13

제12항에 있어서, 상기 연관성 상관 거리는,

상기 제1 브랜치 및 제2 브랜치에서 각각 생성된 두 개의 배치 행렬의 내적을 통해 연관성 상관 행렬(affinity correlation matrix; A)을 획득하고, 상기 연관성 상관 행렬로 상기 불변 항 및 중복 감소 항을 구성해 산출하는 것을 특징으로 하는, 자기-지도 학습 기반의 다중 객체 추적 방법.

청구항 14

제10항에 있어서, 상기 단계 (2)에서는,

확률적 경사 최적화기(stochastic gradient optimizer; SGD)를 적용해 학습 중 평균 손실을 역전파하되, 정지-기울기(stop-gradient)를 사용해 상기 제2 브랜치의 인코더 네트워크의 매개변수를 업데이트하지 않는 것을 특징으로 하는, 자기-지도 학습 기반의 다중 객체 추적 방법.

청구항 15

제9항에 있어서, 상기 단계 (3)에서는,

복수의 추적 객체로 구성된 추적 세트와 복수의 검출 객체로 구성된 검출 세트를 각각 구성하고, 상기 추적 세트와 검출 세트의 모든 요소를 상기 단계 (2)에서 학습된 인코더 네트워크에 각각 입력해 획득한 복수의 추적 특징 벡터와 검출 특징 벡터의 쌍별 유사도를 기반으로 연관성 상관 행렬(A)을 추정하여 최종 검출-추적 객체 쌍을 설정하는 것을 특징으로 하는, 자기-지도 학습 기반의 다중 객체 추적 방법.

청구항 16

제15항에 있어서, 상기 단계 (3)에서는,

데이터 연결 및 헝가리안 매칭을 통해 상기 설정한 최종 검출-추적 객체 쌍의 검출 객체를 결합해 추적 객체 상태를 업데이트하는 것을 특징으로 하는, 자기-지도 학습 기반의 다중 객체 추적 방법.

발명의 설명

기술 분야

본 발명은 다중 객체 추적 장치 및 방법에 관한 것으로서, 보다 구체적으로는 자기-지도 학습 기반의 다중 객체 추적 장치 및 방법에 관한 것이다.

[0001]

배경 기술

- [0002] 객체 추적은 주로 단일 객체 추적 (single-object tracking, SOT)과 다중 객체 추적 (multi-object tracking, MOT)으로 나눌 수 있다. 영상 감시, 자율 주행 차, 로봇 내비게이션 등 실제 적용 분야에서는 여러 물체를 동시에 추적할 수 있는 MOT가 하나의 물체만 추적하는 SOT보다 더 많은 관심을 받고 있다.
- [0004] MOT에서 가장 일반적인 접근 방식인 검출 별 추적 패러다임은 크게 두 가지 성능에 의존한다. 첫 번째는 물체 검출 성능이다. 추적이 후속 추적 작업 중에 끊어지거나 잘못 연결되는 것을 방지하려면 모든 프레임에서 객체를 정확하게 감지해야 한다. 최근 CNN을 기반으로 한 고성능의 다양한 객체 검출기가 도입되었으며, 잘못된 객체 검출로 인한 추적 성능 저하가 어느 정도 해결되었다. 그러나 객체 검출기는 물체 가림이나 카메라 흔들림으로 인해 잘못된 객체를 감지하거나 객체를 놓칠 수 있다. 두 번째는 데이터 연결 성능이다. 잘못된 물체 검출로 인한 MOT의 부정확성을 보상하기 위해 데이터 연결은 이전에 획득한 궤적과 새로운 검출을 연결할 수 있다. 온라인 MOT에서 실시간 데이터 연결은 오프라인 MOT보다 더 강조되고 있다.
- [0006] 실시간 데이터 연결의 핵심 문제는 검출 객체와 추적 객체 간의 최적 연결을 결정하는 것이다. 가장 대표적인 데이터 연결 방법은 이분 할당(bipartite assignment)과 헝가리안 기반(Hungarian-based) 접근이다. 이러한 방법은 기존 궤적 노드와 새로운 검출 노드로 구성된 그래프 세트 간의 연관성 행렬(affinity matrix)로 가중치를 모델링한다. 그래프 세트의 노드 간 일치는 연관성 행렬의 가중치에 따라 결정된다. 동적 프로그래밍은 검출과 트랙렛(tracklet) 사이의 최단 경로를 찾아 일치시키는 데 사용될 수 있다. 최소 비용 흐름(min-cost flow) 및 조건부 임의 필드(conditional random field)는 데이터 연결을 그래프로 처리한다. 이 접근 방식은 검출 또는 트랙렛을 그래프의 노드로 표현하고 흐름 모델(flow model) 또는 레이블은 에지 강도(edge strength)를 예측하고 강도가 높은 노드는 연결된다.
- [0008] MOT 데이터 연관에 적용된 딥 러닝 기술은 최근 최첨단 성능을 달성했다. 삼 네트워크(Siamese Network)와 수정된 triplet 및 quadruplet network는 동일한 네트워크를 검출 객체 및 추적 객체에 적용하고 출력 특성의 차이를 기반으로 유사성을 계산했다. 이러한 종래 방법에서는, 공유 네트워크의 구조가 복잡할수록 매칭 성능이 높아지지만, 실시간 온라인 추적은 더 어려워진다. 삼 네트워크 기반 데이터 연관은 객체 간의 신뢰할 수 있는 쌍별 상호 작용을 인코딩할 수 있지만, 한 장면에서 여러 객체와 상태가 존재하는 고차원 정보를 고려할 때는 적합하지 않다.
- [0010] 한편, 자기-지도 학습(Self-Supervised Learning; SSL)은 최소한의 데이터만으로 스스로 규칙을 찾아 분석하는 기술로, 라벨링을 하나씩 붙여 학습 데이터로 사용하던 기존의 지도 학습 방식의 한계를 해결하기 위한 기술이다. 그러나 아직까지 온라인 다중 객체 추정에 자기-지도 학습이 적용된 바 없다.
- [0012] 한편, 본 발명과 관련된 선행기술로, 공개특허 제10-2021-0111417호(발명의 명칭: 삼 네트워크를 활용하는 강인한 다중 객체 검출 장치 및 방법, 공개일자: 2021년 09월 13일) 등이 개시된 바 있다.

발명의 내용

해결하려는 과제

- [0013] 본 발명은 기존에 제안된 방법들의 상기와 같은 문제점들을 해결하기 위해 제안된 것으로서, 자기-지도 학습을 온라인 다중 객체 추정에 적용한 최초의 시도이며, 객체의 포즈 변화를 고려한 학습을 위해 PN-GAN(pose normalized GAN)과 이미지 증강을 적용하고, PN-GAN을 통해 학습 데이터가 부족한 문제를 해결해 적은 수의 데이터로도 더 나은 결과를 얻을 수 있는 MOT 모델을 설계할 수 있으며, 새로운 손실 함수 ACD(Affinity Correlation Distance)를 제안하여 인코더 네트워크의 학습 시 더 많은 구별 특징 벡터를 생성할 수 있고, end-to-end 방식을 사용하여 학습할 수 있으며, 분류, 클러스터링 및 회귀에 대한 추가 미세 조정이 필요하지 않기 때문에 추가적인 전이 학습 없이 바로 적용이 가능하고, 다른 최신 방법들에 비해 적은 메모리를 차지하며 빠르

고 높은 추적 성능을 제공할 수 있는, 자기-지도 학습 기반의 다중 객체 추적 장치 및 방법을 제공하는 것을 그 목적으로 한다.

과제의 해결 수단

- [0014] 상기한 목적을 달성하기 위한 본 발명의 특징에 따른 자기-지도 학습 기반의 다중 객체 추적 장치는,
- [0015] 다중 객체 추적 장치로서,
- [0016] 사전 학습된 포즈 정규화-생성적 적대 네트워크(pose normalized GAN(Generative Adversarial Network); PN-GAN)를 사용해, 입력 이미지로부터 포즈 변화를 생성해 복수의 포즈 이미지를 획득하며, 상기 입력 이미지와 포즈 이미지에 이미지 증강(image augmentation)을 적용해 증강 이미지를 생성하는 데이터 증강부;
- [0017] 상기 증강 이미지 중 양의 쌍(positive pairs)을 사용해, 다중 객체 추적(Multi-Object Tracking; MOT)을 위한 대조적 구조(contrastive structure)의 인코더 네트워크를 자기-지도 학습(Self-Supervised Learning; SSL) 방식으로 학습하는 자기-지도 학습부; 및
- [0018] 상기 자기-지도 학습부에서 학습된 인코더 네트워크를 사용해 추적 객체와 검출 객체 간의 유사성을 추정하고 객체 추적을 수행하는 객체 추적부를 포함하는 것을 그 구성상의 특징으로 한다.
- [0020] 바람직하게는, 상기 자기-지도 학습부는,
- [0021] 대조적 구조의 두 개의 인코더 네트워크를 포함하여 SSL-MOT를 구성하며, 상기 SSL-MOT는 상기 두 개의 인코더 네트워크 각각을 학습하는 제1 브랜치 및 제2 브랜치로 구성될 수 있다.
- [0023] 더욱 바람직하게는,
- [0024] 상기 제1 브랜치의 인코더 네트워크와 상기 제2 브랜치의 인코더 네트워크는 가중치를 공유할 수 있다.
- [0026] 더욱 바람직하게는, 상기 자기-지도 학습부는,
- [0027] 상기 두 개의 증강 이미지를 상기 제1 브랜치 및 제2 브랜치에 각각 입력해 상기 SSL-MOT를 학습하되, 상기 두 개의 증강 이미지 쌍의 출력 배치 행렬에서 계산된 교차 상관(cross-correlation)을 불변 항(invariance term) 및 중복 감소 항(redundancy reduction term)과 결합해 상기 두 개의 증강 이미지 간의 유사성을 계산하는 연관성 상관 거리(Affinity Correlation Distance; ACD)를 사용하는 손실 함수를 이용해 상기 SSL-MOT를 학습을 할 수 있다.
- [0029] 더더욱 바람직하게는, 상기 연관성 상관 거리는,
- [0030] 상기 제1 브랜치 및 제2 브랜치에서 각각 생성된 두 개의 배치 행렬의 내적을 통해 연관성 상관 행렬(affinity correlation matrix; A)을 획득하고, 상기 연관성 상관 행렬로 상기 불변 항 및 중복 감소 항을 구성해 산출할 수 있다.
- [0032] 더욱 바람직하게는, 상기 자기-지도 학습부는,
- [0033] 확률적 경사 최적화기(stochastic gradient optimizer; SGD)를 적용해 학습 중 평균 손실을 역전파하되, 정지-기울기(stop-gradient)를 사용해 상기 제2 브랜치의 인코더 네트워크의 매개변수를 업데이트하지 않을 수 있다.
- [0035] 바람직하게는, 상기 객체 추적부는,
- [0036] 복수의 추적 객체로 구성된 추적 세트와 복수의 검출 객체로 구성된 검출 세트를 각각 구성하고, 상기 추적 세트와 검출 세트의 모든 요소를 상기 자기-지도 학습부에 의해 학습된 인코더 네트워크에 각각 입력해 획득한 복

수의 추적 특징 벡터와 검출 특징 벡터의 쌍별 유사도를 기반으로 연관성 상관 행렬(A)을 추정하여 최종 검출-추적 객체 쌍을 설정할 수 있다.

- [0038] 더욱 바람직하게는, 상기 객체 추적부는,
- [0039] 데이터 연결 및 헵가리안 매칭을 통해 상기 설정한 최종 검출-추적 객체 쌍의 검출 객체를 결합해 추적 객체 상태를 업데이트할 수 있다.
- [0041] 또한, 상기한 목적을 달성하기 위한 본 발명의 특징에 따른 자기-지도 학습 기반의 다중 객체 추적 방법은,
- [0042] 다중 객체 추적 방법으로서,
- [0043] (1) 사전 학습된 포즈 정규화-생성적 적대 네트워크(pose normalized GAN(Generative Adversarial Network); PN-GAN)를 사용해, 입력 이미지로부터 포즈 변화를 생성해 복수의 포즈 이미지를 획득하며, 상기 입력 이미지와 포즈 이미지에 이미지 증강(image augmentation)을 적용해 증강 이미지를 생성하는 데이터 증강 단계;
- [0044] (2) 상기 증강 이미지 중 양의 쌍(positive pairs)을 사용해, 다중 객체 추적(Multi-Object Tracking; MOT)을 위한 대조적 구조(contrastive structure)의 인코더 네트워크를 자기-지도 학습(Self-Supervised Learning; SSL) 방식으로 학습하는 자기-지도 학습 단계; 및
- [0045] (3) 상기 단계 (2)에서 학습된 인코더 네트워크를 사용해 추적 객체와 검출 객체 간의 유사성을 추정하고 객체 추적을 수행하는 객체 추적 단계를 포함하는 것을 그 구성상의 특징으로 한다.
- [0047] 바람직하게는, 상기 단계 (2)에서는,
- [0048] 대조적 구조의 두 개의 인코더 네트워크를 포함하여 SSL-MOT를 구성하며, 상기 SSL-MOT는 상기 두 개의 인코더 네트워크 각각을 학습하는 제1 브랜치 및 제2 브랜치로 구성될 수 있다.
- [0050] 더욱 바람직하게는,
- [0051] 상기 제1 브랜치의 인코더 네트워크와 상기 제2 브랜치의 인코더 네트워크는 가중치를 공유할 수 있다.
- [0053] 더욱 바람직하게는, 상기 단계 (2)에서는,
- [0054] 상기 두 개의 증강 이미지를 상기 제1 브랜치 및 제2 브랜치에 각각 입력해 상기 SSL-MOT를 학습하되, 상기 두 개의 증강 이미지 쌍의 출력 배치 행렬에서 계산된 교차 상관(cross-correlation)을 불변 항(invariance term) 및 중복 감소 항(redundancy reduction term)과 결합해 상기 두 개의 증강 이미지 간의 유사성을 계산하는 연관성 상관 거리(Affinity Correlation Distance; ACD)를 사용하는 손실 함수를 이용해 상기 SSL-MOT를 학습할 수 있다.
- [0056] 더더욱 바람직하게는, 상기 연관성 상관 거리는,
- [0057] 상기 제1 브랜치 및 제2 브랜치에서 각각 생성된 두 개의 배치 행렬의 내적을 통해 연관성 상관 행렬(affinity correlation matrix; A)을 획득하고, 상기 연관성 상관 행렬로 상기 불변 항 및 중복 감소 항을 구성해 산출할 수 있다.
- [0059] 더욱 바람직하게는, 상기 단계 (2)에서는,
- [0060] 확률적 경사 최적화기(stochastic gradient optimizer; SGD)를 적용해 학습 중 평균 손실을 역전파하되, 정지-기울기(stop-gradient)를 사용해 상기 제2 브랜치의 인코더 네트워크의 매개변수를 업데이트하지 않을 수 있다.

[0062] 바람직하게는, 상기 단계 (3)에서는,

[0063] 복수의 추적 객체로 구성된 추적 세트와 복수의 검출 객체로 구성된 검출 세트를 각각 구성하고, 상기 추적 세트와 검출 세트의 모든 요소를 상기 단계 (2)에서 학습된 인코더 네트워크에 각각 입력해 획득한 복수의 추적 특징 벡터와 검출 특징 벡터의 쌍별 유사도를 기반으로 연관성 상관 행렬(A)을 추정하여 최종 검출-추적 객체 쌍을 설정할 수 있다.

[0065] 더욱 바람직하게는, 상기 단계 (3)에서는,

[0066] 데이터 연결 및 헵가리안 매칭을 통해 상기 설정한 최종 검출-추적 객체 쌍의 검출 객체를 결합해 추적 객체 상태를 업데이트할 수 있다.

발명의 효과

[0067] 본 발명에서 제안하고 있는 자기-지도 학습 기반의 다중 객체 추적 장치 및 방법에 따르면, 자기-지도 학습을 온라인 다중 객체 추정에 적용한 최초의 시도이며, 객체의 포즈 변화를 고려한 학습을 위해 PN-GAN(pose normalized GAN)과 이미지 증강을 적용하고, PN-GAN을 통해 학습 데이터가 부족한 문제를 해결해 적은 수의 데이터로도 더 나은 결과를 얻을 수 있는 MOT 모델을 설계할 수 있으며, 새로운 손실 함수 ACD(Affinity Correlation Distance)를 제안하여 인코더 네트워크의 학습 시 더 많은 구별 특징 벡터를 생성할 수 있고, end-to-end 방식을 사용하여 학습할 수 있으며, 분류, 클러스터링 및 회귀에 대한 추가 미세 조정이 필요하지 않기 때문에 추가적인 전이 학습 없이 바로 적용이 가능하고, 다른 최신 방법들에 비해 적은 메모리를 차지하며 빠르고 높은 추적 성능을 제공할 수 있다.

도면의 간단한 설명

[0068] 도 1은 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 장치의 구성을 도시한 도면.

도 2는 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 장치에서 수행되는 학습 과정을 나타낸 도면.

도 3은 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 장치의 학습 알고리즘을 나타낸 도면.

도 4는 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 장치에서 수행되는 테스트 과정을 나타낸 도면.

도 5는 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 방법의 흐름을 도시한 도면.

도 6은 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 장치 및 방법에서, 데이터 증강 방식에 따른 MOT 수행 결과를 비교하여 나타낸 도면.

도 7은 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 장치 및 방법에서, 손실 함수에 사용한 유사성 메트릭에 따른 수행 결과를 나타낸 도면.

도 8은 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 장치 및 방법에서, 손실 함수에 따른 인코더 네트워크의 출력 벡터의 판별력을 시각화하여 나타낸 도면.

도 9는 SSL 접근법에 따른 MOT 수행 결과를 나타낸 도면.

도 10은 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 장치 및 방법과 최신 방법들을 MOT 16 데이터셋을 사용해 실험한 결과를 비교하여 나타낸 도면.

도 11은 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 장치 및 방법과 최신 방법들을 MOT 17 데이터셋을 사용해 실험한 결과를 비교하여 나타낸 도면.

도 12는 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 장치 및 방법과 최신 방법들을 MOT 20 데이터셋을 사용해 실험한 결과를 비교하여 나타낸 도면.

발명을 실시하기 위한 구체적인 내용

- [0069] 이하, 첨부된 도면을 참조하여 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 본 발명을 용이하게 실시할 수 있도록 바람직한 실시예를 상세히 설명한다. 다만, 본 발명의 바람직한 실시예를 상세하게 설명함에 있어, 관련된 공지 기능 또는 구성에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략한다. 또한, 유사한 기능 및 작용을 하는 부분에 대해서는 도면 전체에 걸쳐 동일한 부호를 사용한다.
- [0071] 덧붙여, 명세서 전체에서, 어떤 부분이 다른 부분과 ‘연결’되어 있다고 할 때, 이는 ‘직접적으로 연결’되어 있는 경우뿐만 아니라, 그 중간에 다른 소자를 사이에 두고 ‘간접적으로 연결’되어 있는 경우도 포함한다. 또한, 어떤 구성요소를 ‘포함’한다는 것은, 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있다는 것을 의미한다.
- [0073] 도 1은 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 장치(100)의 구성을 도시한 도면이다. 도 1에 도시된 바와 같이, 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 장치(100)는, 데이터 증강부(110), 자기-지도 학습부(120) 및 객체 추적부(130)를 포함하여 구성될 수 있다. 여기서, 데이터 증강부(110)와 자기-지도 학습부(120)는 학습 과정, 객체 추적부(130)는 테스트 또는 실제 온라인 MOT에 적용해 객체를 추적하는 과정을 각각 처리할 수 있다.
- [0075] 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 장치(100) 및 방법은, 자기-지도 학습을 온라인 다중 객체 추정에 적용한 최초의 시도로서, 다음과 같은 특징을 갖는다: (1) 객체의 포즈 변화를 고려한 학습을 위해 PN-GAN(pose normalized GAN)과 이미지 증강을 적용한다. (2) PN-GAN을 통해 학습 데이터가 부족한 문제를 해결해 적은 수의 데이터로도 더 나은 결과를 얻을 수 있는 MOT 모델을 설계할 수 있다. (3) 새로운 손실 함수 ACD(Affinity Correlation Distance)를 제안하여 인코더 네트워크의 학습 시 더 많은 구별 특징 벡터를 생성할 수 있다. (4) end-to-end 방식을 사용하여 학습할 수 있으며, 분류, 클러스터링 및 회귀에 대한 추가 미세 조정이 필요하지 않기 때문에 추가적인 전이 학습 없이 바로 적용이 가능하다. (5) 다른 최신 방법들에 비해 적은 메모리를 차지하며 빠르고 높은 추적 성능을 제공할 수 있다.
- [0077] 이하에서는, 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 장치(100)의 각 구성요소에 대해 상세히 설명하기 전에, 본 발명과 관련된 선행 연구인 자기-지도 학습에 대해 설명하도록 한다.
- [0079] 일반적으로 대조 학습의 한 유형인 SimCLR(Chen et al., 2020)은 삼 네트워크와 유사한 구조를 갖지만 양수 및 음수 샘플 쌍을 정의하여 다른 방식으로 손실 함수를 계산하는 데 사용된다. BYOL(Grill et al., 2020)은 SimCLR(Chen et al., 2020)에서 사용된 음수 쌍을 사용하여 이미지의 대상 네트워크 표현을 예측하는 대신 동일한 이미지를 다르게 증강하여 온라인 네트워크를 학습했다. Barlow Twins(Zbotar et al., 2021)는 입력 이미지에서 두 개의 왜곡된 샘플을 두 개의 동일한 네트워크에 공급하고 두 출력 간의 교차-상관 행렬(cross-correlation matrix)을 측정했다. 이 방법은 교차-상관 행렬을 단위행렬에 최대한 가깝게 만들어 축소 표현(collapse representation)을 자연스럽게 피하는 목적 함수를 적용한다. 그러나 축소 표현을 피하기 위해서 여전히 큰 배치 사이즈가 필요하며 출력 차원이 4096으로 크기 때문에 실시간 연결에 적합하지 않다. SimSiam(Chen and He, 2021)은 음의 쌍 또는 운동량 인코더를 사용하지 않고, 대칭 손실을 사용해 이미지의 두 뷰의 유사성을 직접 최대화한다. 정지-기울기(stop-gradient) 작업을 사용하여 축소 표현(collapse representation)을 방지할 수 있고 일반 배치 크기로 작동하며 대규모 배치 교육에 의존하지 않는다. SimSiam(Chen and He, 2021)과 유사하게 SCARF(Bahri et al., 2021)는 공유 가중치를 가진 두 개의 인코더로 구성되지만 두 브랜치는 공유 가중치로 사전 학습된 헤드를 가지고 있다. 그러나 전술한 바와 같은 기존 SSL 방식들은 엄격한 end-to-end 학습을 적용하지 않고 인코더를 학습한 후 클러스터링, 분류, 그리고 회귀 등에 적용하기 위해 전이 학습을 수행해야 한다.

- [0081] 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 장치(100)에서는, SSL의 기본 구조는 Chen and He(2021)(Chen, X.; and He, K. 2021. Exploring Simple Siamese Representation Learning. In CVPR.)에 의해 언급된 구조를 사용하되, 사람의 자세 변화에 강하고 end-to-end 학습 및 테스트가 가능하도록, 내부 학습 절차를 MOT 최적화를 위해 재구성하였다. 이는 SSL을 온라인 MOT에 최초로 적용한 것이다.
- [0083] 이하에서는, 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 장치(100)를 구성하는 각 구성 요소에 대해 상세히 설명하도록 한다.
- [0085] 데이터 증강부(110)는, 사전 학습된 포즈 정규화-생성적 적대 네트워크(pose normalized GAN(Generative Adversarial Network); PN-GAN)를 사용해, 입력 이미지로부터 포즈 변화를 생성해 복수의 포즈 이미지를 획득하며, 입력 이미지와 포즈 이미지에 이미지 증강(image augmentation)을 적용해 증강 이미지를 생성할 수 있다.
- [0087] 도 2는 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 장치(100)에서 수행되는 학습 과정을 나타낸 도면이다. 도 2에 도시된 바와 같이, 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 장치(100)는, 데이터 증강부(110)에서 생성한 증강 이미지를 사용해, 자기-지도 학습부(120)에서 SSL-MOT를 학습할 수 있다.
- [0089] SSL에서는, 이미지의 여러 가지 다른 증강을 수행한 후 양수 또는 음수 쌍을 사용해 네트워크 학습을 수행한다. SSL에 사용되는 이미지 증강 기술은, 임의 자르기(random cropping), 크기 조정(resizing), 뒤집기(fliping), 색상 지터링(color jittering) 및 가우시안 블러(Gaussian blur) 등을 포함할 수 있다. 그러나 사람 추적에서는 포즈 변화가 빈번하게 발생하고, 단순한 이미지 확대로는 정확한 이미지 매칭이 어렵다.
- [0091] 따라서 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 장치(100)의 데이터 증강부(110)는, 도 1의 좌측 “Data Augmentation”에 도시된 바와 같이, 이미지 증강을 위한 전처리로 PN-GAN(Qian et al., 2018)을 채택하였다. PN-GAN에서는 하나의 보행자 이미지가 제공될 때 원하는 다중 포즈에 따라 동일한 ID와 다른 포즈를 가진 합성 이미지가 생성된다. 데이터 증강부(110)는, 입력 이미지 x를 PN-GAN에 공급하여 생성된 포즈 이미지를 얻고, 이미지 증강을 포즈 이미지에 무작위로 적용할 수 있다. 그러나 PN-GAN은 전신이 영상에 포함된 경우에만 새로운 포즈를 생성할 수 있으므로, OpenPose(Cao et al., 2021)를 전처리로 사용하여 포즈가 감지된 경우에만 PN-GAN을 적용할 수 있다. 즉, 데이터 증강부(110)에서는, 입력 이미지에서 전신 포즈가 감지된 경우에는 PN-GAN을 사용해 복수의 포즈 변화를 생성하고, 생성된 포즈 이미지에 이미지 증강을 적용하는 반면에, 전신 포즈가 감지되지 않은 경우에는 이미지 증강만 적용할 수 있다.
- [0093] 자기-지도 학습부(120)는, 증강 이미지 중 양의 쌍(positive pairs)을 사용해, 다중 객체 추적(Multi-Object Tracking; MOT)을 위한 대조적 구조(contrastive structure)의 인코더 네트워크를 자기-지도 학습(Self-Supervised Learning; SSL) 방식으로 학습할 수 있다.
- [0095] SSL-MOT 네트워크는 학습 인코더에 대해 두 개의 증강 이미지를 사용할 수 있다. SSL-MOT 네트워크는 대조적 구조의 두 개의 인코더 네트워크로 구성되며, 각각의 인코더 네트워크를 학습하는 두 개의 브랜치(branch)로 구성될 수 있다.
- [0097] 보다 구체적으로, 도 2의 우측 “Training of SSL-MOT”에 도시된 바와 같이, 자기-지도 학습부(120)는, 대조적 구조의 두 개의 인코더 네트워크를 포함하여 SSL-MOT를 구성하며, SSL-MOT는 두 개의 인코더 네트워크 각각을

학습하는 제1 브랜치 및 제2 브랜치로 구성될 수 있다. 또한, 제1 브랜치의 인코더 네트워크와 제2 브랜치의 인코더 네트워크는 가중치를 공유할 수 있다.

[0099] 여기서, 제1 브랜치는, 입력된 증강 이미지로부터 표현 벡터를 출력하는 인코더 네트워크(Encoder); 및 인코더 네트워크에서 출력되는 표현 벡터를 새로운 출력 벡터로 출력하는 예측 네트워크(Predictor)를 포함하여 구성될 수 있다. 즉, 제1 브랜치의 첫 번째 네트워크는 인코더 네트워크 $f(\cdot)$ 를 사용하여 증강된 입력 이미지에서 표현 벡터를 출력할 수 있다. 그런 다음 인코더 네트워크 $f(\cdot)$ 는 ResNet50을 백본 네트워크로 하여 프로젝션(Projection) MLP와 결합할 수 있다. 다른 네트워크는 예측 헤드 $h(\cdot)$ 로 한 뷰를 변환해 다른 뷰와 일치시키는 역할을 할 수 있다.

[0101] 제2 브랜치는, 제1 브랜치의 인코더 네트워크와 가중치를 공유하는 인코더 네트워크(Encoder)를 포함하여 구성될 수 있다. 즉, 제2 브랜치는 단 하나의 인코더 네트워크 $f(\cdot)$ 로 구성될 수 있으며, 프로젝션 헤드는 MLP를 사용하며 테스트용이 아닌 학습용으로만 사용될 수 있다.

[0103] SSL-MOT 학습 후 전이 학습을 위한 추가 학습이 필요하지 않도록, SSL-MOT는 두 입력 이미지(입력된 증강 이미지)의 유사성을 확인할 수 있는 연관성 행렬(affinity matrix)을 출력하도록 종단(end-to-end) 간 학습이 필요하다.

[0105] 도 3은 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 장치(100)의 학습 알고리즘을 나타낸 도면이다. 즉, 도 3은 N개의 라벨이 없는 이미지를 사용해 SSL-MOT를 학습하는 과정을 나타낸 것으로, 이하에서는 도 2 및 도 3을 참조하여 자기-지도 학습부(120)의 SSL-MOT 학습 과정을 상세히 설명하도록 한다.

[0107] 배치 크기가 n 이고 미니 배치 B가 $\{x_i\}_{i=1}^n$ 으로 로드된다고 가정한다. 도 2의 좌측 “Data Augmentation”과 같이 이미지 x_i 에 이미지 증강이 적용되고, 증강 이미지 중 하나는 추적 객체(tracker) tr_i 이고 다른 하나는 검출 객체(detection) d_i 라고 가정한다. 각 증강 이미지는 가중치를 공유하는 인코더 네트워크 $f(\cdot)$ 에 공급되며, 그 출력은 $z_{tr_i}=f(tr_i)$ 및 $z_{d_i}=f(d_i)$ 로 표시될 수 있다. 제1 브랜치의 z_{tr_i} 는 새로운 M차원 출력 벡터 p_{tr_i} 를 생성하기 위해 예측 헤드에 다시 공급되며, $p_{tr_i}=h(f(tr_i))$ 로 표현될 수 있다. 제2 브랜치에서 인코더는 동일한 M차원 벡터 z_d 를 출력할 수 있다. 그러나 예측 헤드에 공급되지 않고 사용될 수 있다.

[0109] 이전의 SSL 기반 방법(Chen et al., 2020; Caron et al., 2020; Grill et al., 2020; Zbotar et al., 2021; Chen and He, 2021)은 수백만 개의 이미지에서 표현을 학습했으며, 목적(분류, 클러스터링 및 회귀)에 의존적이어서 일부 지도 작업(supervised tasks)에서 전이 학습과 미세 조정이 필요했다. 이러한 접근 방식은 엄밀히 말하면 종단 간 학습(end-to-end learning)이 아니며, 인코더가 학습된 후 목적에 따라 개별적으로 다시 전이 학습을 수행해야 한다.

[0111] 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 장치(100)의 자기-지도 학습부(120)는, 전이 학습 없이 분류, 클러스터링, 회귀 등에 적용 가능하도록 종단 간 학습을 하기 위해, 새로운 손실 함수를 사용하였다.

[0113] 즉, 자기-지도 학습부(120)는, 두 개의 증강 이미지를 제1 브랜치 및 제2 브랜치에 각각 입력해 SSL-MOT를 학습하되, 두 개의 증강 이미지 쌍의 출력 배치 행렬에서 계산된 교차 상관(cross-correlation)을 불변 항(invariance term) 및 중복 감소 항(redundancy reduction term)과 결합해 두 개의 증강 이미지 간의 유사성을

계산하는 연관성 상관 거리(Affinity Correlation Distance; ACD)를 사용하는 손실 함수를 이용해 SSL-MOT를 학습을 할 수 있다. 여기서, 연관성 상관 거리(ACD)는, 제1 브랜치 및 제2 브랜치에서 각각 생성된 두 개의 배치 행렬의 내적을 통해 연관성 상관 행렬(affinity correlation matrix; A)을 획득하고, 연관성 상관 행렬로 불변 항 및 중복 감소 항을 구성해 산출할 수 있다. 이하에서는, SSL-MOT의 학습을 위해 사용하는 손실 함수에 대해 상세히 설명하도록 한다.

[0115] SSL에서 주로 사용되는 손실 함수는 평균 제곱 오차(MSE)(Grill et al., 2020), L1 손실, 코사인 유사도(Chen and He, 2021), 교차 엔트로피, 교차 상관 행렬(Zbontar, et al., et al., 2021) 등이다. 이러한 손실의 목표는 출력 벡터가 잘 클러스터링 되도록 SSL이 학습된 표현을 생성하도록 학습하는 것이다. 따라서 추가적인 전이 학습 없이 한 번의 중단 간 학습(one stage end-to-end learning)을 통해 두 이미지 간의 유사도를 계산할 수 있는 손실 함수가 필요하다.

[0117] 본 발명에서는, SSL에서 일반적으로 사용되는 손실 함수 대신, 이미지 쌍의 출력 배치 행렬에서 계산된 교차 상관(cross-correlation)을 불변 항(invariance term) 및 중복 감소 항(redundancy reduction term)과 결합하는 새로운 연관성 상관 거리 (Affinity Correlation Distance; ACD)를 제안한다. 그러나 Zbontar et al. (2021)과는 달리, 여기서는 특징 차원 간의 표현 상관관계가 아니라, 배치 샘플 간의 상관관계를 관찰했다. 즉, 배치 내 유사한 샘플 쌍 간의 상관 값은 크고 인접 샘플에 대한 상관 값은 작다.

[0119] 보다 구체적으로, 먼저 n 개의 미니 배치에서 제1 브랜치로부터 배치 행렬 $\mathbf{p}_{tr} \in \mathbb{R}^{n \times M}$ 을 생성하고, 제2 브랜치에서는 배치 행렬 $\mathbf{z}_d \in \mathbb{R}^{n \times M}$ 을 생성할 수 있다. 그런 다음, 수학적 식 1과 같이, 두 배치 행렬의 내적을 통해 연관성 상관 행렬(affinity correlation matrix) A를 얻을 수 있다.

수학적 식 1

[0121] $\mathbf{A} = \mathcal{N}(\mathbf{p}_{tr} \cdot \mathbf{z}_d^T) \in \mathbb{R}^{n \times n}$

[0123] 여기서, N 은 L-2 정규화를 나타낸다. 연관성 상관 행렬 A는 다음 수학적 식 2의 ACD를 계산하는 데 사용된다.

수학적 식 2

[0125] $ACD(\mathbf{p}_{tr}, \mathbf{z}_d) = \sum_i (1 - A_{ii})^2 + \lambda \sum_i \sum_{i \neq j} A_{ij}^2$

[0127] 여기서, λ 는 손실의 첫 번째 및 두 번째 항의 중요도를 상쇄하는 가중치이고 i 와 j 는 배치 샘플의 인덱스이다. 수학적 식 2 우변의 첫 번째 항인 불변 항에서 동일한 샘플 쌍은 0의 값을 생성하려고 시도하고, 두 번째 항인 중복 감소 항에서는 다른 샘플에 대해 0에 가까운 값을 갖도록 시도한다.

[0129] ACD를 적용함으로써, 목적에 따라 표현 학습(representation learning)과 별도의 추가적인 전이 학습을 적용하는 문제를 해결할 수 있다. 학습 후, 학습된 인코더에 의해 생성된 출력 벡터의 내적만을 사용하여 추적 객체와 검출 객체의 연관성 점수(affinity score)를 쉽게 계산할 수 있다.

[0131] 두 인코더 네트워크가 가중치를 공유하는 SSL-MOT는 두 관측치가 동일한 출력을 생성하는 모델링 불변성을 보장해야 한다. 따라서 Chen and He(2021) 및 Grill et al. (2020)와 유사하게, 증강 이미지는 서로 다른 브랜치에 공급되고 $ACD(p_d, z_{tr})$ 가 계산될 수 있다. 또한, 대칭 손실(symmetry loss)은 다음 수학적 식 3과 같이 두 ACD를 사용하여 계산될 수 있다.

수학적 식 3

$$\mathcal{L} = \frac{1}{2} \{ ACD(p_{tr}, z_d) + ACD(p_d, z_{tr}) \}$$

[0133]

[0135] 한편, 자기-지도 학습부(120)는, 확률적 경사 최적화기(stochastic gradient optimizer; SGD)를 적용해 모든 학습 단계에서 학습 중 평균 손실을 역전파하되, 정지-기울기(stop-gradient)를 사용해 제2 브랜치의 인코더 네트워크의 매개변수를 업데이트하지 않을 수 있다. 이는 제1 브랜치의 인코더 및 예측 네트워크의 매개변수 θ 만 업데이트하고 제2 브랜치의 인코더 매개변수는 업데이트하지 않는데, 이는 다음 수학적 식 4와 같이 나타낼 수 있다.

수학적 식 4

$$\theta_f, \theta_h \leftarrow \text{SGD}(\theta, \nabla \mathcal{L}, \eta)$$

[0137]

[0139] 여기서, η 는 학습률이다. 정지-기울기를 사용하면 출력 벡터를 생성할 수 없다. 증강된 추적 객체 tr 및 감지 객체 d는 제1 브랜치와 정지 기울기가 적용된 제2 브랜치에 교대로 공급되므로, 수학적 식 3은 다음 수학적 식 5와 같이 변경될 수 있다.

수학적 식 5

$$\mathcal{L} = \frac{1}{2} \{ ACD(p_{tr}, \text{stopgrad}(z_d)) + ACD(p_d, \text{stopgrad}(z_{tr})) \}$$

[0141]

[0143] 객체 추적부(130)는, 자기-지도 학습부(120)에서 학습된 인코더 네트워크를 사용해 추적 객체와 검출 객체 간의 유사성을 추정하고 객체 추적을 수행할 수 있다. 객체 추적부(130)는, 복수의 추적 객체로 구성된 추적 세트와 복수의 검출 객체로 구성된 검출 세트를 각각 구성하고, 추적 세트와 검출 세트의 모든 요소를 자기-지도 학습부(120)에 의해 학습된 인코더 네트워크에 각각 입력해 획득한 복수의 추적 특징 벡터와 검출 특징 벡터의 쌍별 유사도를 기반으로 연관성 상관 행렬(A)을 추정하여 최종 검출-추적 객체 쌍을 설정할 수 있다. 보다 구체적으로는, 객체 추적부(130)는, 데이터 연결 및 헵가리안 매칭을 통해 설정한 최종 검출-추적 객체 쌍의 검출 객체를 결합해 추적 객체 상태를 업데이트할 수 있다.

[0145] 도 4는 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 장치(100)에서 수행되는 테스트 과정을 나타낸 도면이다. 이하에서는, 도 4를 참조하여 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객

체 추적 장치(100)의 객체 추적부(130)에 대해 상세히 설명하도록 한다.

[0147] 먼저, 학습된 SSL-MOT를 온라인 객체 추적에 적용할 수 있다. 보다 구체적으로, 자기-지도 학습부(120)의 SSL-MOT 학습 후, 실제 추적 중 추적 객체와 검출 객체 간의 연관성 매트릭스(affinity matrix)를 학습된 SSL-MOT를 사용하여 생성할 수 있다. t 번째 프레임($t > 1$)에서 H 추적 객체로 구성된 추적 세트를 $\mathbf{T}^{t-1} = \{tr_1^{t-1}, tr_2^{t-1}, \dots, tr_h^{t-1}\}$, 그리고 G 검출 객체로 구성된 검출 세트를 $\mathbf{D}^t = \{d_1^t, d_2^t, \dots, d_g^t\}$ 라고 하자. 개별 추적 객체 및 검출 객체는 각각 $tr_h^t = [x_h^t, y_h^t, w_h^t, h_h^t]$, $d_g^t = [x_g^t, y_g^t, w_g^t, h_g^t]$ 와 같이 중심 위치(x, y), 너비(w), 높이(h) 요소로 구성된 벡터로 표현될 수 있다. 모든 프레임에서 검출 세트 $\mathbf{D}$ 의 각 요소 d_g 는 학습된 인코더 f 에 입력되고, L2-정규화는 새로운 d -차원 특징 벡터 z_{dg} 를 생성하기 위해 출력에 적용될 수 있다. 변환된 검출 객체의 세트 $\mathbb{Z}_d$ 는 세트 $\mathbf{D}$ 의 모든 요소에 대해 동일한 연산을 통해 획득될 수 있다. 마찬가지로, 동일한 인코더가 추적 객체의 각 요소 tr_i 에 적용되어 새로운 특징 벡터 z_{trh} 를 얻을 수 있다. 변환된 추적 객체 $\mathbb{Z}_{tr}$ 의 세트도 세트 $\mathbf{T}$ 의 모든 요소에 대해 동일한 작업을 통해 얻을 수 있다.

[0149] SSL-MOT의 목표는 온라인 MOT에서 $\mathbf{T}$ 추적 객체와 $\mathbf{D}$ 검출 객체로부터 추출된 특징들의 쌍별 유사도(pair-wise similarity)를 기반으로 $H \times G$ 연관성 상관 행렬 $\mathbf{A} = \{z_{hg}\}^{|\mathbf{T}^{t-1}| \times |\mathbf{D}^t|}$ 을 추정하는 것이다. 추적 객체와 검출 객체 사이에 연관성 상관 행렬 $\mathbf{A}$ 를 생성하기 위해 집합 $\mathbb{Z}_{tr}$ 을 $(H \times d)$ 행렬로 변환하고, 집합 $\mathbb{Z}_d$ 를 $(G \times d)^T$ 로 전치할 수 있다. 그런 다음 수학적 6과 같이 두 행렬 사이의 내적을 취한다.

수학적 6

$$\mathbf{A} = \exp \left\{ \frac{1}{\tau} \cdot (\mathbb{Z}_{tr} \cdot \mathbb{Z}_d^T) \right\}$$

[0153] 여기서, τ 는 출력 값 간의 차이를 넓히는 데 사용되는 상수 값이다.

[0155] $\mathbf{A}$ 의 요소로서 z_{hg} 는 검출 객체 d_g^t 가 추적 객체 tr_h^{t-1} 과 연관되어 있으면 큰 값을 갖고, 두 객체가 연관되지 않으면 작은 값을 가질 수 있다. 연관성 상관 행렬에서 연관성 점수가 임계값 ρ_1 보다 작으면 해당 쌍의 연관성 점수를 0으로 필터링하여 일치 속도(matching speed)를 향상시킬 수 있다. 비용 함수(cost function) c 의 경우, 다음 수학적 7과 같이 가중치 합을 사용하여 외형 공간(appearance space)에서 연관성 점수를 상대적 L1 중심 거리 $\text{Dis}(tr_h^{t-1}, d_g^t)$ 와 결합할 수 있다.

수학적 7

$$c(tr_h^{t-1}, d_g^t) = \alpha \cdot z_{hg} + \beta \cdot \text{Dis}(tr_h^{t-1}, d_g^t)$$

[0159] 여기서, α 와 β 는 각각 0.5와 0.5인 가중치를 나타낼 수 있다. 비용 함수 c 는 헝가리안 매칭(Hungarian

matching)에 사용될 수 있다.

[0161] 첫 번째 프레임을 제외한 모든 프레임에서 추적 객체와 검출 객체는 제안된 SSL-MOT 기반 데이터 연결 및 형가리안 매칭을 통해 새 연결을 업데이트하여 궤적을 생성할 수 있다. 데이터 연관을 통해 최종 검출-추적 객체 쌍이 결정되면, 이전 추적 객체와 현재 검출 객체를 결합하여 다음 프레임에 대해 $tr_h^t = \delta \cdot tr_h^{t-1} + (1 - \delta) \cdot d_g^t$ 와 같이 추적 객체 상태를 업데이트할 수 있다. 여기서, δ 는 추적 객체의 상태를 제어하는 조정 가능한 매개변수이며 일반적으로 0.5로 설정될 수 있다.

[0163] 온라인 MOT 매칭 규칙을 다음과 같이 정의한다.

[0165] 첫째, 검출 객체가 추적 객체와 일치하지 않으면 이 검출 객체를 새로운 잠재적 추적 객체로 할당하고, 잠재적 추적 객체가 임계값 p_2 번 이상 일치하면 참 추적 객체로 새로 할당하며($T^t \leftarrow T^{t-1} \cup (d_g^t, h+1)$), 그렇지 않으면 거짓 추적 객체로 인식되어 추적 객체 풀에서 제거될 수 있다.

[0167] 둘째, 추적 객체가 검출 객체와 일치하지 않으면 교합(occlusion)에 의해 일시적으로 사라질 수 있으므로 즉시 제거하지 않고 임계값 p_2 프레임 동안 추적 객체의 상태를 유지하면서 관찰할 수 있다. 그러나 p_2 프레임 이상 일치하는 항목이 없으면 누락된 추적 객체로 간주되어 추적 세트에서 제거할 수 있다($T^t \leftarrow T^{t-1} \setminus tr_h^{t-1}$). p_2 의 값은 데이터 세트의 특성에 따라 조정할 수 있다.

[0169] 도 5는 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 방법의 흐름을 도시한 도면이다. 도 5에 도시된 바와 같이, 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 방법은, 사전 학습된 포즈 정규화-생성적 적대 네트워크(pose normalized GAN(Generative Adversarial Network); PN-GAN)를 사용해, 입력 이미지로부터 포즈 변화를 생성해 복수의 포즈 이미지를 획득하며, 입력 이미지와 포즈 이미지에 이미지 증강(image augmentation)을 적용해 증강 이미지를 생성하는 데이터 증강 단계(S100), 증강 이미지 중 양의 쌍(positive pairs)을 사용해, 다중 객체 추적(Multi-Object Tracking; MOT)을 위한 대조적 구조(contrastive structure)의 인코더 네트워크를 자기-지도 학습(Self-Supervised Learning; SSL) 방식으로 학습하는 자기-지도 학습 단계(S200) 및 단계 S200에서 학습된 인코더 네트워크를 사용해 추적 객체와 검출 객체 간의 유사성을 추정하고 객체 추적을 수행하는 객체 추적 단계(S300)를 포함하여 구현될 수 있다.

[0171] 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 방법은, 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 장치(100)에서 각 단계가 수행될 수 있으며, 메모리 및 프로세서를 포함한 하드웨어에서 기록되는 소프트웨어로 구성될 수 있다. 예를 들어, 개인용 컴퓨터, 노트북 컴퓨터, 서버 컴퓨터, PDA, 스마트폰, 태블릿 PC 등에 저장 및 구현될 수 있다.

[0173] 각각의 단계들과 관련된 상세한 내용들은, 앞서 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 장치(100)와 관련하여 충분히 설명되었으므로, 상세한 설명은 생략하기로 한다.

[0175] **실험**

[0177] MOT 벤치마크 데이터 세트 (MOT-BD)를 사용하여 본 발명에서 제안하는 SSL-MOT의 객체 추적 성능을 평가하였다. 본 발명의 타당성을 입증하기 위해 애블레이션 연구(ablation study) 및 최첨단 방법과의 비교를 포함하여 다양

한 실험을 수행하였다.

[0179] 구현 세부 정보

[0181] 본 발명에서 제안된 SSL-MOT를 Pytorch 및 Pytorch-Geometric을 사용하여 구현하였으며, Nvidia GeForce RTX 2080ti GPU를 사용하는 Intel Core i9-9900k CPU 클러스터에서 학습 및 테스트했다. 본 발명의 네트워크 구성과 학습 설정은 다음과 같다.

[0183] • OpenPose는 다단계 CNN으로 구성된다. 첫 번째 단계는 부분 연관성 필드(part affinity fields)를 예측하고, 두 번째 단계는 신뢰 맵(confidence map)을 예측한다. PN-GAN 생성기는 ResNet-50을 기반으로 설계되었으며, 주어진 입력 이미지에 대해 4개의 연속적인 표준 포즈를 생성하는 인코더-디코더 네트워크로 구성하였다.

[0185] • 이미지 증강 파이프라인은 5가지 이미지 변환(가우시안 블러링, 수평 뒤집기, 원근감, 컷아웃, 밝기 지터링)으로 구성되며 특정 확률로 무작위로 적용된다.

[0187] • Encoder f 는 백본으로 ResNet-50과 3개의 추가 레이어로 구성된 MLP로 구성된다. 배치 크기는 512이고 각 완전 연결(fc) 계층에 대해 배치 정규화가 사용되었다. 옵티마이저로 확률적 경사 최적화기(SGD)를 사용하였고, 학습률은 $\eta \times \text{BatchSize}/256$ 에, 기본 $\eta=0.05$ 이고 가중치 감소는 1×10^{-4} 로 하였다.

[0189] • 예측 네트워크 h 는 두 개의 fc 레이어로 구성되며 그 출력은 2048-d이다.

[0191] 데이터 세트

[0193] SSL-MOT 학습을 위해 Market-ket1501(Zheng et al., 2015)을 MOT16(Milan et al., 2016) 및 MOT17(MOT 2021) 학습 데이터 세트와 결합했다. MOT20(Dendorfer et al., 2020) 학습 데이터 세트는 심각한 교합 때문에 학습에 포함하지 않았다. Market1501에는 6대의 카메라에서 가져온 12,936개의 학습 이미지가 포함되어 있다. 테스트를 위해서는, MOT 챌린지 데이터 세트, MOT16, MOT17 및 MOT20에 대한 실험을 수행했다. 다른 방법과의 공정한 비교를 위해 MOT Challenge에서 제공하는 공개 탐지를 사용하여 MOT 성능을 측정했다.

[0195] 평가 지표

[0197] 본 발명의 정량적 평가를 위해 multiple object tracking accuracy (MOTA), identity F1 score (IDF1), number of false positives (FP), number of false negatives (FN), number of ID switches (IDsw) 및 Hz(frames per second)를 사용하였다.

[0199] 실험예 1. 애블레이션 연구 (Ablation Study)

[0201] 개별 구성 요소의 중요성을 이해하기 위해 본 발명에서 제안된 SSL-MOT에 대해 일련의 애블레이션 연구를 수행하였다. MOT 평가는 온라인 평가 서버를 통해서만 수행할 수 있고, 테스트 영상의 성능은 4번의 시도로 제한되

어 있기 때문에, 테스트 세트 대신 MOT17 학습 세트의 비디오 시퀀스를 사용하여 애플리케이션 성능을 측정했다. MOT 문헌(Papakis, et al. 2021; Li et al., 2020)의 표준 관행으로, MOT17 학습 세트(MOT17-02, 05, 10, 11)와 테스트 세트(MOT17-04, 09, 13)의 정지 및 움직이는 카메라에서 촬영된 5개 비디오 시퀀스를 사용했습니다.

[0203] 1-1. PN-GAN 이미지 증강의 효과

[0205] 본 발명에서는, 기존의 SSL 방법과 달리 PN-GAN과 이미지 증강을 이용하여 객체 추적에 최적화된 일반화된 SSL-MOT 네트워크를 학습한다. 일반적인 이미지 확대가 아니라, 사람의 포즈를 인식하고 다음 프레임에서 변경될 수 있는 새로운 포즈를 생성한다.

[0207] PN-GAN 기반 데이터 증강의 효과를 입증하기 위해 (1) SSL의 일반적인 이미지 증강, (2) 이미지 증강이 없는 PN-GAN, (3) 이미지 증강을 포함하는 PN-GAN, (4) OpenPose, PN-GAN 및 이미지 증강 (제안된 방법)의 4가지 증강 사례에 대한 비교 실험을 수행했다. 제안된 방법은 보행자의 빈번한 패색과 카메라 시점을 고려하여 임의의 패치만을 선택하고 크기를 조정하고 수평으로 뒤집는 방식으로 이미지 증강을 적용하였다.

[0209] 도 6은 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 장치(100) 및 방법에서, 데이터 증강 방식에 따른 MOT 수행 결과를 비교하여 나타낸 도면이다. 도 6에 도시된 바와 같이, SSL-MOT는 물체의 자세를 예측하여 학습하기 때문에 기본적인 이미지 증강 방법 (1)에 비해 PN-GAN 기반 방법((2), (3) 및 (4))의 전반적인 성능이 향상되었다. 방법 (4)의 경우 포즈 변화뿐만 아니라 이미지 증강도 고려하여 학습하기 때문에, 방법 (3)에 비해 MOTA는 0.8%, IDF1은 3.2% 증가하였다. 방법 (2)가 IDF1과 FP에서 방법 (4)보다 약간 더 나은 성능을 보였지만, 방법 (4)는 MOT에서 중요한 지표인 MOTA에서 3% 이상 더 높은 성능을 보였다. 따라서 이하 실험에서는 PN-GAN과 OpenPose를 결합하고 이미지 증강을 사용하는 방법 (4)를 적용해 학습된 SSL-MOT 모델을 사용한다.

[0211] 1-2. 손실에 대한 유사성 메트릭(similarity metric)의 효율성

[0213] 제안된 유사성 메트릭 ACD의 성능을 측정하기 위해 동일한 네트워크 구조에서 대칭 손실에 사용되는 유사성 메트릭을 변경하면서 이미지 매칭 성능을 측정하였다. 실험 동안 SOTA SSL의 손실 함수에 사용된 세 가지 주요 방법인 MSE(Zbontar et al., 2021), 코사인 유사도(Chen and He, 2021) 및 제안된 ACD의 성능을 비교했다.

[0215] 도 7은 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 장치(100) 및 방법에서, 손실 함수에 사용한 유사성 메트릭에 따른 수행 결과를 나타낸 도면이다. 도 7에 도시된 바와 같이, MSE는 3가지 지표에서 가장 낮은 성능을 보였다. 크기를 고려하지 않고 각도 거리만 고려하는 코사인 유사도와 달리 ACD는 출력 벡터 사이의 각도 및 크기 속성을 유지할 수 있는 내적 결과를 사용하여 최상의 성능을 보였다.

[0217] 인코더는 클래스 내 가변성을 줄이고 클래스 간 가변성을 증가시키도록 학습되며, 인코더 출력 벡터는 ACD의 불변 및 중복 감소 항을 사용하여 클래스 간 판별(inter-class discrimination)을 증가시키는 특성을 갖는다. ACD를 사용하여 학습된 인코더 네트워크의 출력 벡터의 판별력(discrimination power)을 테스트하기 위해 t-SNE를 사용하여 출력 벡터를 시각화했다.

[0219] 도 8은 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 장치(100) 및 방법에서, 손실 함수에 따른 인코더 네트워크의 출력 벡터의 판별력을 시각화하여 나타낸 도면이다. 도 8은 인코더 학습 후 10명의 인물 이미지를 입력하여 t-SNE로 표현한 것으로, (a) MSE, (b) 코사인 유사도, (3) 제안된 ACD를 각각 나타낸다.

도 8에 도시된 바와 같이, 제안된 ACD는 MSE와 코사인 유사도에 비해 클래스 내 및 클래스 간 변동성 측면에서 우수한 판별력을 보였고, 임베딩된 특징 공간 내에서 동일인 간의 유사도를 최대화하였다.

[0221] 1-3. SSL 접근 방식에 따른 MOT 비교

[0223] MOT에 가장 유용한 SSL 구조를 찾기 위해 제안된 SSL-MOT를 MOT17 데이터 세트를 사용하여 테스트하고, 데이터 증강을 사용하여 학습된 BYOL(Grill et al., 2020), SimSiam(Chen and He, 2021), Barlow Twins(Zbontar et al., 2021) 및 SSL-MOT에 대해 4가지 성능 메트릭을 측정하여 비교하였다.

[0225] 도 9는 SSL 접근법에 따른 MOT 수행 결과를 나타낸 도면이다. 도 9에 도시된 바와 같이, 제안된 SSL-MOT는 모든 메트릭에서 최고의 성능을 보였다. SimSiam과 SSL-MOT는 유사한 기본 네트워크 구조를 가지고 있으나, SSL-MOT는 ACD 기반 손실 함수를 통해 SimSiam에 비해 추적 객체 매칭 성능을 향상시킬 수 있었다. 테스트 중 SSL-MOT는 출력 벡터에 대한 내적만 계산하여 정확한 친화도 점수를 추정할 수 있으므로, 일치하는 대상이 많은 경우에도 다른 방법에 비해 더 짧은 데이터 연관 시간과 더 빠른 속도(Hz)를 얻을 수 있다.

[0227] 실험예 2. MOT Challenge 벤치마크 평가

[0229] 제안된 방법의 MOT 성능을 검증하기 위해 MOT16, 17 및 20 테스트 데이터 세트에 대해 SOTA(State of the art) MOT 방법을 사용하여 비교 실험을 수행했다. 비교 방법은, SOTA 온라인 추적 방법 중에서 선택했다.

[0231] MOT16 데이터 세트에 대한 MOT 성능은 다음과 같이 4가지 SOTA 온라인 MOT 접근 방식을 사용하여 측정되었다: (1) BLSTM_MTP_0 (Kim et al., 2021), (2) MFI_TST (Yang et al., 2021), (3) ArTIST-T (Saleh et al., 2021), (4) SiamesRF (Lee et al., 2020).

[0233] 도 10은 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 장치(100) 및 방법과 최신 방법들을 MOT 16 데이터셋을 사용해 실험한 결과를 비교하여 나타낸 도면이다. 도 10에서 보는 바와 같이, 제안한 방법(SSL-MOT)은 모든 방법에서 빠른 추적 속도를 유지하면서 가장 높은 MOTA와 FN 점수를 얻었다.

[0235] MOT17 데이터 세트에서는, 동일한 온라인 기반 MOT 방법에 CenterTrack(Zhou et al., 2020)을 추가해 실험하였다.

[0237] 도 11은 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 장치(100) 및 방법과 최신 방법들을 MOT 17 데이터셋을 사용해 실험한 결과를 비교하여 나타낸 도면이다. 도 11에서 보는 바와 같이, CenterTrack은 MOT17에 대해 우수한 성능을 보였다. 그러나 CenterTrack은 공개검출을 사용하기 때문에 MOT 공개 탐지만 사용하는 다른 방법과 직접 비교할 수 없다. CenterTrack을 제외하고, 제안된 방법은 MOT16 결과와 유사하게 최고의 MOTA 및 FN을 달성했다.

[0239] MOT20 데이터 세트에서는, CenterTrack 대신 TMOH(Stadler and Beyerer, 2021)을 추가해 실험하였다.

[0241] 도 12는 본 발명의 일실시예에 따른 자기-지도 학습 기반의 다중 객체 추적 장치(100) 및 방법과 최신 방법들을 MOT 20 데이터셋을 사용해 실험한 결과를 비교하여 나타낸 도면이다. 도 12에서 보는 바와 같이, TMOH가 온라인 MOT 방식 중 가장 좋은 성능을 보였지만, 이 방식 역시 CenterTrack과 같은 공개검출 방식을 이용하여 처리

속도가 0.6Hz로 매우 느린 단점이 있다. 제안하는 SSL-MOT 방법은 검출 정련을 사용하지 않고 MOT20의 학습 데이터를 학습에서 제외하였기 때문에 낮은 성능을 보였다. 그러나 제안된 방법은 온라인상에서 1.6Hz의 가장 빠른 추적 속도를 보였다.

[0243] 실험을 통해, 제안하는 방법이 SOTA 온라인 MOT 접근 방식 중 추적 정확도와 속도 측면에서 우수한 성능을 발휘함을 확인할 수 있었다. 전술한 바와 같이, 본 발명에서 제안하고 있는 자기-지도 학습 기반의 다중 객체 추적 장치(100) 및 방법에 따르면, 자기-지도 학습을 온라인 다중 객체 추정에 적용한 최초의 시도이며, 객체의 포즈 변화를 고려한 학습을 위해 PN-GAN(pose normalized GAN)과 이미지 증강을 적용하고, PN-GAN을 통해 학습 데이터가 부족한 문제를 해결해 적은 수의 데이터로도 더 나은 결과를 얻을 수 있는 MOT 모델을 설계할 수 있으며, 새로운 손실 함수 ACD(Affinity Correlation Distance)를 제안하여 인코더 네트워크의 학습 시 더 많은 구별 특징 벡터를 생성할 수 있고, end-to-end 방식을 사용하여 학습할 수 있으며, 분류, 클러스터링 및 회귀에 대한 추가 미세 조정이 필요하지 않기 때문에 추가적인 전이 학습 없이 바로 적용이 가능하고, 다른 최신 방법들에 비해 적은 메모리를 차지하며 빠르고 높은 추적 성능을 제공할 수 있다.

[0245] 한편, 본 발명은 다양한 통신 단말기로 구현되는 동작을 수행하기 위한 프로그램 명령을 포함하는 컴퓨터에서 판독 가능한 매체를 포함할 수 있다. 예를 들어, 컴퓨터에서 판독 가능한 매체는, 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media) 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치를 포함할 수 있다.

[0247] 이와 같은 컴퓨터에서 판독 가능한 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 이때, 컴퓨터에서 판독 가능한 매체에 기록되는 프로그램 명령은 본 발명을 구현하기 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 예를 들어, 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해 실행될 수 있는 고급 언어 코드를 포함할 수 있다.

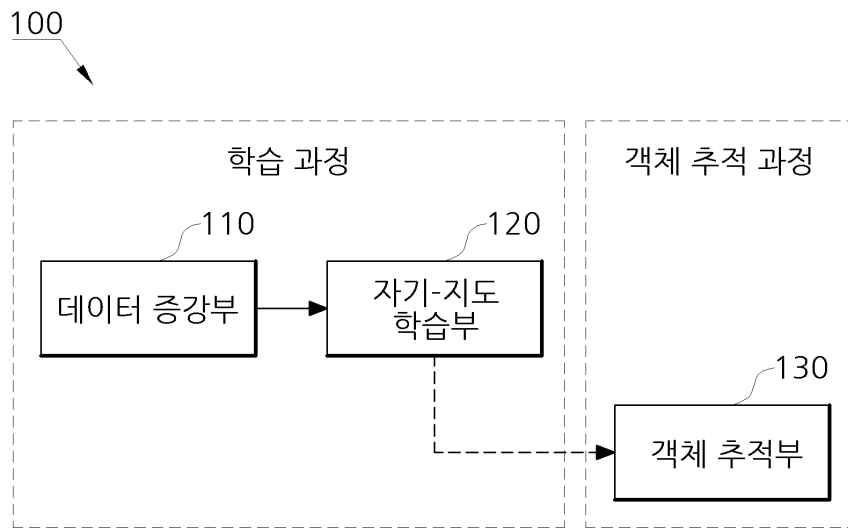
[0249] 이상 설명한 본 발명은 본 발명이 속한 기술분야에서 통상의 지식을 가진 자에 의하여 다양한 변형이나 응용이 가능하며, 본 발명에 따른 기술적 사상의 범위는 아래의 특허청구범위에 의하여 정해져야 할 것이다.

부호의 설명

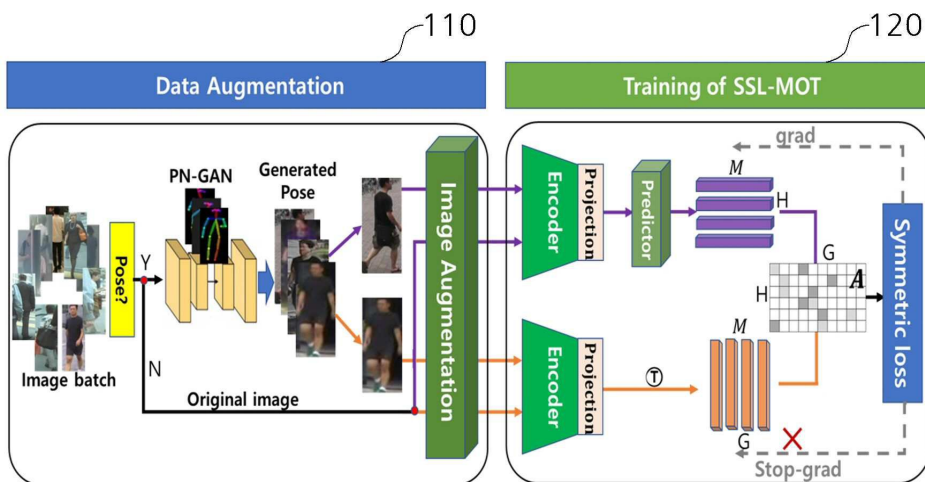
[0250] 100: 다중 객체 추적 장치
110: 데이터 증강부
120: 자기-지도 학습부
130: 객체 추적부
S110: 데이터 증강 단계
S120: 자기-지도 학습 단계
S130: 객체 추적 단계

도면

도면1



도면2



도면3

Algorithm 1: Training of SSL-MOT

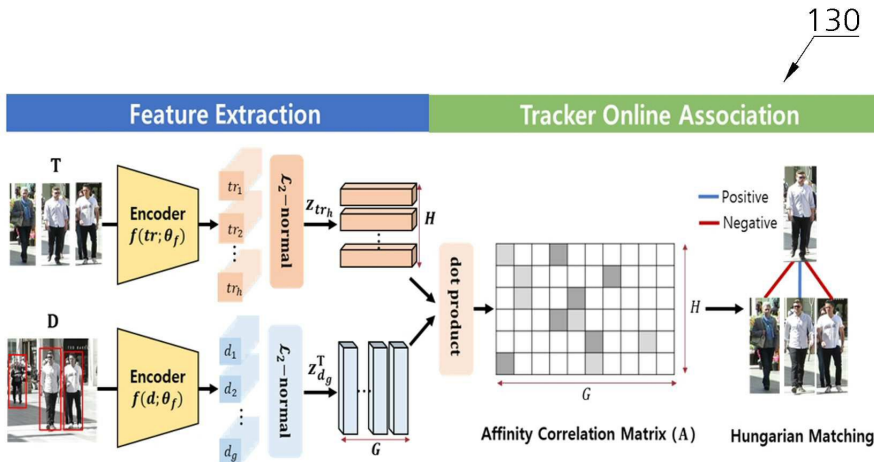
Input: Optimizer: SGD , No. of images: N ,
Total number of optimization steps: K
Parameter: θ_f for f and θ_h for h
Output: Trained Siamese encoder (f_θ)

```

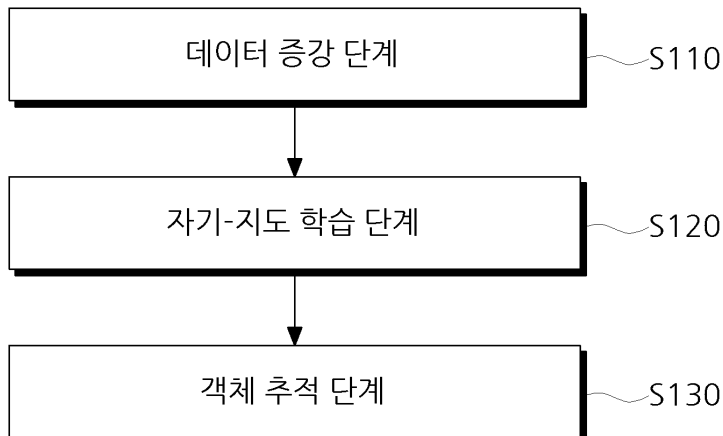
1: for  $k=1, \dots, K$  do
2:    $\mathcal{B} \leftarrow \{x_i \sim D\}_{i=1}^n$ 
3:   for  $x_i \in \mathcal{B}$  do
4:      $tr_i \leftarrow aug(x_i)$  // image augmentation
5:      $d_i \leftarrow aug(x_i)$ 
6:      $z_{tr_i} \leftarrow stopgrad(f(tr_i))$  // stop gradient
7:      $z_{d_i} \leftarrow stopgrad(f(d_i))$ 
8:      $p_{tr_i} \leftarrow h(z_{tr_i})$  //compute prediction of  $tr_i$  and  $d_i$ 
9:      $p_{d_i} \leftarrow h(z_{d_i})$ 
10:  end
11:  // compute affinity correlation distance
12:   $ACD(p_{tr}, z_d)^k, ACD(p_d, z_{tr})^k$ 
13:  // compute the symmetric loss
14:   $\mathcal{L}_k = \frac{1}{2} \{ACD(p_{tr}, z_d)^k + ACD(p_d, z_{tr})^k\}$ 
15:   $\delta\theta_k \leftarrow \partial_{\theta} \mathcal{L}_k$  //compute a gradient
16:   $\theta_f, \theta_h \leftarrow SGD(\theta_k, \delta\theta_k, \eta)$  //update parameters
17: end

```

도면4



도면5



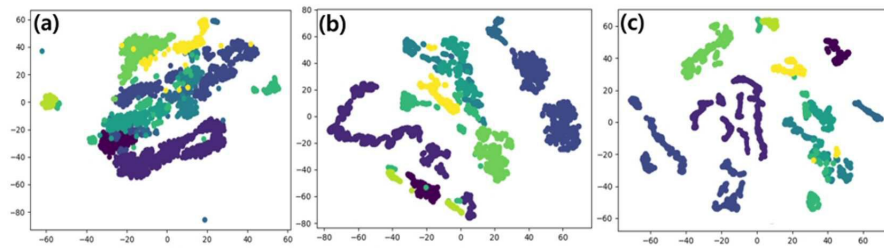
도면6

Augmentation	MOTA↑	IDF1↑	FP↓
(1) w/ Image Aug.	59.6%	55.1%	922
(2) PN-GAN+ w/o Image Aug.	65.9%	68.3%	749
(3) PN-GAN+ w/ Image Aug.	68.1%	62.3%	758
(4) PN-GAN+OpenPose+w/ Image Aug.	68.9%	65.5%	763

도면7

Method	MOTA↑	IDF1↑	FP↓
MSE	60.3	41.0	1,247
Cosine similarity	65.7	49.5	964
Proposed ACD	68.9	65.5	763

도면8



도면9

Method	MOTA↑	IDF1↑	FP↓	Hz
BYOL	51.3	36.3	3,040	2.18
SimSiam	40.4	21.1	6,448	10.01
Barlow Twins	62.1	51.3	1,378	0.6
SSL-MOT	68.9	65.5	763	13.2

도면10

Trackers	MOTA↑	IDF1↑	FP↓	FN↓	IDsw↓	Hz
BLSTM_MTP_O	48.3	53.5	9,792	83,707	435	21.0
MFI_TST	59.9	58.7	3,660	68,923	616	0.7
ArTIST-T	56.6	57.8	3,532	75,031	-	4.5
SiamesRF	57.9	41.7	8,196	66,538	2,051	11.1
SSL-MOT	62.1	45.2	4,760	63,238	1,022	13.7

도면11

Trackers	MOTA↑	IDF1↑	FP↓	FN↓	IDsw↓	Hz
CenterTrack	61.5	59.6	14,076	200,672	2,583	17
BLSTM_MTP_O	51.5	54.9	29,616	241,619	2,566	20.1
MFI_TST	60.1	58.8	13,503	209,475	2,065	2.2
ArTIST-T	56.7	57.5	12,353	230,437	-	4.5
SiamesRF	55.8	46.7	12,822	233,573	3,174	12.4
SSL-MOT	61.5	55.0	16,841	197,485	2,957	12.9

도면12

Trackers	MOTA↑	IDF1↑	FP↓	FN↓	IDsw↓	Hz
TMOH	60.1	61.2	38,043	165,899	2,342	0.6
MFI_TST	59.2	59.1	36,402	172,785	1,914	0.5
ArTIST-C	53.6	51.0	7,765	230,576	1,531	1.0
SSL-MOT	48.1	37.5	70,716	186,779	11,247	1.6