



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0074231
(43) 공개일자 2024년05월28일

(51) 국제특허분류(Int. Cl.)

G06T 7/246 (2017.01) G06T 7/215 (2017.01)

G06V 10/44 (2022.01) G06V 10/74 (2022.01)

(52) CPC특허분류

G06T 7/246 (2017.01)

G06T 7/215 (2017.01)

(21) 출원번호 10-2022-0156062

(22) 출원일자 2022년11월21일

심사청구일자 2022년11월21일

(71) 출원인

인하대학교 산학협력단

인천광역시 미추홀구 인하로 100(용현동, 인하대 학교)

(72) 발명자

배승환

인천광역시 미추홀구 인하로 100(용현동)

유용상

인천광역시 미추홀구 인하로67번길 24-12, 201호 (용현동)

이성호

인천광역시 미추홀구 경인로 67(송의동, The 미추홀퍼스트)

(74) 대리인

양정보

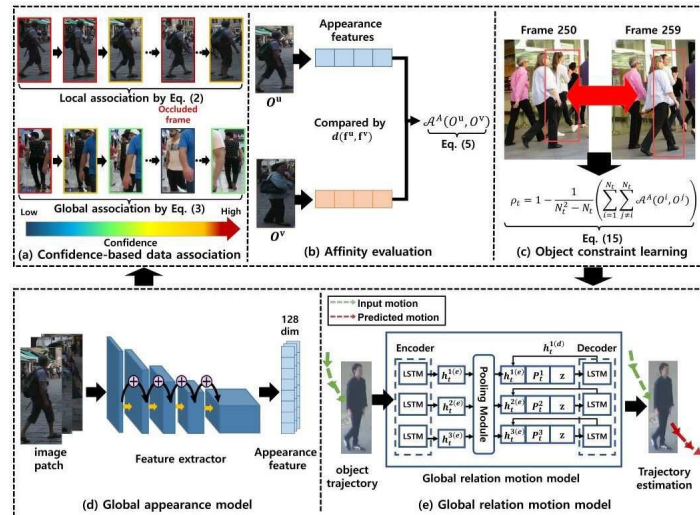
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 객체 전체 모델들 및 모델 상태 기반 온라인 제한 학습을 통한 효율적인 다중 객체 추적 방법 및 장치

(57) 요약

본 개시는 객체 전체 모델들 및 모델 상태 기반 온라인 제한 학습을 통한 효율적인 다중 객체 추적 방법 및 장치를 제공하며, 글로벌 외관 모델을 이용하여, 연속 프레임들에서 감지되는 객체들의 외관 특징들을 검출하고, 글로벌 상대 모션 모델을 이용하여, 감지되는 객체들의 미래 궤적들을 예측하고, 외관 특징들과 미래 궤적들을 기반으로 감지되는 객체들을 구별하여 연계시키도록 구성될 수 있다. 본 개시에 따르면, 글로벌 외관 모델 또는 글로벌 상대 모션 모델 중 적어도 하나는, 학습 시 미리 정해진 제약 조건을 기반으로 업데이트될 수 있다.

대표도



(52) CPC특허분류

G06V 10/44 (2023.08)

G06V 10/761 (2023.08)

G06T 2207/30241 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711154594
과제번호	2022R1C1C1009208
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	우수신진연구
연구과제명	[Ezbaro] 대규모 확장 가능한 On-Vision AI를 위한 E3-AI 네트워크와 협업 학습 연

구

기 여 율	70/100
과제수행기관명	인하대학교 산학협력단
연구기간	2022.03.01 ~ 2023.02.28

이 발명을 지원한 국가연구개발사업

과제고유번호	1711160586
과제번호	2022-0-00448-001
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	방송통신산업기술개발사업
연구과제명	[Ezbaro] 인간처럼 회상이 가능한 인공 신경망 지속학습 플랫폼 개발(1차년도)
기 여 율	10/100
과제수행기관명	인하대학교 산학협력단
연구기간	2022.04.01 ~ 2022.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711152759
과제번호	2020-0-01389-003
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	인공지능융합연구센터지원사업(국고)
연구과제명	[Ezbaro][지자체] 인공지능융합연구센터지원(3차년도)
기 여 율	10/100
과제수행기관명	인하대학교 산학협력단
연구기간	2022.01.01 ~ 2022.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711174118
과제번호	RS-2022-00155915
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신인재양성(교육운영수익)
연구과제명	[Ezbaro][정부+민간] 인공지능융합혁신인재양성(1차년도)
기 여 율	10/100
과제수행기관명	인하대학교 산학협력단
연구기간	2022.07.01 ~ 2022.12.31

명세서

청구범위

청구항 1

컴퓨터 장치의 다중 객체 추적 방법에 있어서,
 글로벌 외관 모델을 이용하여, 연속 프레임들에서 감지되는 객체들의 외관 특징들을 검출하는 단계;
 글로벌 상대 모션 모델을 이용하여, 상기 감지되는 객체들의 미래 궤적들을 예측하는 단계; 및
 상기 외관 특징들과 상기 미래 궤적들을 기반으로 상기 감지되는 객체들을 구별하여 연계시키는 단계
 를 포함하고,
 상기 글로벌 외관 모델 또는 상기 글로벌 상대 모션 모델 중 적어도 하나는,
 학습 시 미리 정해진 제약 조건을 기반으로 업데이트되는,
 다중 객체 추적 방법.

청구항 2

제 1 항에 있어서,
 상기 글로벌 외관 모델은, 학습 시,
 훈련 데이터셋에서 감지되는 객체들을 아이디(ID)들로 정의하고,
 타겟인 앵커 객체, 상기 앵커 객체와 동일한 아이디의 포지티브 객체, 및 상기 앵커 객체와 다른 아이디의 네거티브 객체를 선택하고,
 상기 앵커 객체와 상기 포지티브 객체의 외관 특징들에 따른 포지티브 객체 쌍들, 및 상기 앵커 객체와 상기 네거티브 객체의 외관 특징들에 따른 네거티브 객체 쌍들을 정의하고,
 트리플렛(triplet) 손실을 기반으로, 상기 포지티브 객체 쌍들 사이의 특징 거리를 최소화하고, 상기 네거티브 객체 쌍들 사이의 특징 거리를 최대화하는,
 다중 객체 추적 방법.

청구항 3

제 1 항에 있어서,
 상기 글로벌 상대 모션 모델은,
 훈련 데이터셋에서 감지되는 객체들의 각각의 과거 궤적 및 상기 객체들 사이의 상대 모션을 기반으로 상기 객체들의 각각의 미래 궤적을 예측하도록, 학습되는,
 다중 객체 추적 방법.

청구항 4

제 3 항에 있어서,
 상기 글로벌 상대 모션 모델은,
 상기 과거 궤적 및 상기 상대 모션을 기반으로 상기 객체들의 각각의 미래 궤적을 생성하도록 구성되는 생성기;

및

상기 예측된 미래 궤적이 상기 훈련 데이터셋에서 확인되는 실측 궤적과 일치하는 지의 여부를 판별하도록 구성되는 판별기

를 포함하는,

다중 객체 추적 방법.

청구항 5

제 1 항에 있어서,

새 객체가 나타나는 경우, 상기 글로벌 외관 모델 및 상기 글로벌 상대 모션 모델이 업데이트되고,

한 객체에 대해 상이한 프레임들에서 검출되는 외관 특징들의 유사도가 임계값 이하인 경우, 상기 글로벌 외관 모델이 업데이트되고,

현재 프레임에서 예측되는 미래 궤적들의 수가 미리 정해진 수 미만이면, 상기 글로벌 상대 모션 모델이 업데이트되는,

다중 객체 추적 방법.

청구항 6

다중 객체 추적 장치에 있어서,

메모리; 및

상기 메모리와 연결되고, 상기 메모리에 저장된 적어도 하나의 명령을 실행하도록 구성되는 프로세서

를 포함하고,

상기 프로세서는,

글로벌 외관 모델을 이용하여, 연속 프레임들에서 감지되는 객체들의 외관 특징들을 검출하고,

글로벌 상대 모션 모델을 이용하여, 상기 감지되는 객체들의 미래 궤적들을 예측하고,

상기 외관 특징들과 상기 미래 궤적들을 기반으로 상기 감지되는 객체들을 구별하여 연계시키도록 구성되고,

상기 글로벌 외관 모델 또는 상기 글로벌 상대 모션 모델 중 적어도 하나는,

학습 시 미리 정해진 제약 조건을 기반으로 업데이트되는,

다중 객체 추적 장치.

청구항 7

제 6 항에 있어서,

상기 글로벌 외관 모델은, 학습 시,

훈련 데이터셋에서 감지되는 객체들을 아이디(ID)들로 정의하고,

타겟인 앵커 객체, 상기 앵커 객체와 동일한 아이디의 포지티브 객체, 및 상기 앵커 객체와 다른 아이디의 네거티브 객체를 선택하고,

상기 앵커 객체와 상기 포지티브 객체의 외관 특징들에 따른 포지티브 객체 쌍들, 및 상기 앵커 객체와 상기 네거티브 객체의 외관 특징들에 따른 네거티브 객체 쌍들을 정의하고,

트리플릿(triplet) 손실을 기반으로, 상기 포지티브 객체 쌍들 사이의 특징 거리를 최소화하고, 상기 네거티브

객체 쌍들 사이의 특징 거리를 최대화하는,
다중 객체 추적 장치.

청구항 8

제 6 항에 있어서,
상기 글로벌 상대 모션 모델은,
훈련 데이터셋에서 감지되는 객체들의 각각의 과거 궤적 및 상기 객체들 사이의 상대 모션을 기반으로 상기 객체들의 각각의 미래 궤적을 예측하도록, 학습되는,
다중 객체 추적 장치.

청구항 9

제 8 항에 있어서,
상기 글로벌 상대 모션 모델은,
상기 과거 궤적 및 상기 상대 모션을 기반으로 상기 객체들의 각각의 미래 궤적을 생성하도록 구성되는 생성기;
및
상기 예측된 미래 궤적이 상기 훈련 데이터셋에서 확인되는 실측 궤적과 일치하는 지의 여부를 판별하도록 구성되는 판별기
를 포함하는,
다중 객체 추적 장치.

청구항 10

제 6 항에 있어서,
상기 프로세서는,
새 객체가 나타나는 경우, 상기 글로벌 외관 모델 및 상기 글로벌 상대 모션 모델을 업데이트하고,
한 객체에 대해 상이한 프레임들에서 검출되는 외관 특징들의 유사도가 임계값 이하인 경우, 상기 글로벌 외관 모델을 업데이트하고,
현재 프레임에서 예측되는 미래 궤적들의 수가 미리 정해진 수 미만이면, 상기 글로벌 상대 모션 모델을 업데이트하도록 구성되는,
다중 객체 추적 장치.

발명의 설명

기술 분야

[0001] 본 개시는 객체 전체 모델들(global objects models) 및 모델 상태 기반 온라인 제한 학습을 통한 효율적인 다중 객체 추적 방법 및 장치에 관한 것이다.

배경 기술

[0002] 다중 객체 추적(multi-object tracking; MOT)은 다수의 객체들의 상태들을 찾는 다음 연속 프레임들의 사이에서 객체들을 정확하게 연계시키는 문제이다. 다중 객체 추적은 지난 몇 년 동안 감시 시스템, 자율 주행 및 스포츠

분석과 같은 많은 응용 분야들에 적용되었다. 추적-감지(tracking-by-detection) 패러다임은 일반적으로 다중 객체 추적의 문제를 해결하기 위해 사용되며 인상적인 성능 향상을 달성했다. 감지 응답들이 각 프레임에서 감지기에 의해 제공되면, 감지들(detections)이 프레임들의 전체에 걸쳐 연결 또는 연계된다. 그러나, 추적 정확도와 속도 사이의 균형에 따른 효율적인 다중 객체 추적을 달성하기에는 어려움이 있다.

발명의 내용

해결하려는 과제

[0003] 본 개시는 객체 전체 모델들 및 모델 상태 기반 온라인 제한 학습을 통한 효율적인 다중 객체 추적 방법 및 장치를 제공한다.

과제의 해결 수단

[0004] 본 개시의 컴퓨터 장치의 다중 객체 추적 방법은, 글로벌 외관 모델을 이용하여, 연속 프레임들에서 감지되는 객체들의 외관 특징들을 검출하는 단계, 글로벌 상대 모션 모델을 이용하여, 상기 감지되는 객체들의 미래 궤적들을 예측하는 단계, 및 상기 외관 특징들과 상기 미래 궤적들을 기반으로 상기 감지되는 객체들을 구별하여 연계시키는 단계를 포함하고, 상기 글로벌 외관 모델 또는 상기 글로벌 상대 모션 모델 중 적어도 하나는, 학습 시 미리 정해진 제약 조건을 기반으로 업데이트될 수 있다.

[0005] 본 개시의 다중 객체 추적 장치는, 메모리, 및 상기 메모리와 연결되고, 상기 메모리에 저장된 적어도 하나의 명령을 실행하도록 구성되는 프로세서를 포함하고, 상기 프로세서는, 글로벌 외관 모델을 이용하여, 연속 프레임들에서 감지되는 객체들의 외관 특징들을 검출하고, 글로벌 상대 모션 모델을 이용하여, 상기 감지되는 객체들의 미래 궤적들을 예측하고, 상기 외관 특징들과 상기 미래 궤적들을 기반으로 상기 감지되는 객체들을 구별하여 연계시키도록 구성되고, 상기 글로벌 외관 모델 또는 상기 글로벌 상대 모션 모델 중 적어도 하나는, 학습 시 미리 정해진 제약 조건을 기반으로 업데이트될 수 있다.

발명의 효과

[0006] 본 개시에 따르면, 객체 전체 모델들을 이용하여 효율적인 다중 객체 추적이 가능하다. 구체적으로, 글로벌 외관 모델에 의해 객체들에 대한 높은 식별성을 확보하고, 글로벌 상대 모션 모델에 의해 각 객체의 정확한 궤적 추정이 가능하므로, 다중 객체 추적의 추적 정확도와 속도를 함께 향상시킬 수 있다. 또한, 객체 전체 모델들에 대한 객체 제한 학습을 통해, 온라인에서의 객체 전체 모델들을 학습하는 계산 비용이 완화될 수 있다.

도면의 간단한 설명

[0007] 도 1은 본 개시에 따른 컴퓨터 장치를 개략적으로 도시하는 도면이다.

도 2는 본 개시에 따른 컴퓨터 장치에 의한 객체 전체 모델들 및 모델 상태 기반 온라인 제한 학습을 통한 효율적인 다중 객체 추적의 전체 프레임워크를 도시하는 도면이다.

도 3은 본 개시에 따른 글로벌 외관 모델의 심층 특징 추출기의 아키텍처를 도시하는 도면이다.

도 4는 본 개시에 따른 글로벌 외관 모델에서 트리플릿 손실을 사용하여 객체 특징 거리들을 변경하는 예를 도시하는 도면이다.

도 5는 본 개시에 따른 글로벌 외관 모델에서의 하드 포지티브 및 네거티브 샘플들에 대한 예들을 제공하는 도면이다.

도 6은 본 개시에 따른 글로벌 상대 모션 모델의 아키텍처를 도시하는 도면이다.

도 7은 본 개시에 따른 객체 제한 학습을 설명하기 위한 예들을 제공하는 도면이다.

도 8 및 도 9는 본 개시에 따른 알고리즘들을 나타내는 도면들이다.

도 10은 본 개시에 따른 컴퓨터 장치의 방법을 개략적으로 도시하는 도면이다.

발명을 실시하기 위한 구체적인 내용

[0008] 이하, 본 개시의 다양한 실시예들이 첨부된 도면을 참조하여 설명된다.

- [0009] 이하에서, 본 개시는 객체 전체 모델들 및 모델 상태 기반 온라인 제한 학습을 통한 효율적인 다중 객체 추적 방법 및 장치를 제공한다.
- [0010] 본 개시는 효율적인 다중 객체 추적을 위해 낮은 학습 및 추론 복잡성을 유지하면서 추적되는 서로 다른 객체들을 정확하게 구별할 수 있는 객체 전체 모델들, 바꿔 말해, 글로벌 객체 모델들을 제안한다. 글로벌 모델들로서, 먼저 대조 학습을 사용하여 글로벌 외관 모델을 설계한다. 구체적으로, 가벼운 ConvNet에서 추적된 객체들에 대한 높은 수준의 시맨틱 객체 특징들을 추출한다. 그런 다음, 식별들을 고려하여 포지티브(positive)(즉, 샘플) 및 네거티브(negative)(즉, 다른) 객체 특징 쌍들을 정의한다. 트리플릿 손실(triplet loss)을 기반으로, 포지티브 객체 쌍들 사이의 특징 거리를 최소화하고 네거티브 객체 쌍들의 특징 거리를 최대화한다. 또한, 이전 모션들에서 객체의 미래 궤적을 예측하는 글로벌 상대 모션 모델을 제시한다. 글로벌 상대 모션 모델의 주요 모션 큐들로서 각 객체의 과거 궤적 및 추적된 객체들 간의 상대 모션을 사용합니다. 모션 예측 결과들을 개선하기 위해 글로벌 상대 모션 모델은 주로 생성기와 판별기로 구성되고 적대적 학습을 사용하여 그들을 훈련한다. 글로벌 외관 및 모션 모델들이 효과적인 객체 모델들이지만, 각 프레임에서 이러한 모델들을 온라인으로 학습하면 다중 객체 추적의 복잡성이 증가한다.
- [0011] 따라서, 본 개시는 이러한 모델들을 적응적으로 업데이트하는 객체 제한 학습을 제안한다. 주요 아이디어는 해당 모델들의 식별 가능성이 추적된 객체들을 구별하기에 불충분할 때 이러한 모델들을 업데이트하는 것이다. 이를 달성하기 위해, 모델 식별 가능성의 객체 제약 조건을 정의하고 제약 조건과 불일치가 발생하면 객체 모델들을 업데이트한다. 그 대가로, 온라인 학습 복잡성도 줄일 수 있다. 신뢰도 기반 데이터 연계(association)를 위해 글로벌 외관 및 모션 모델들과 객체 제약 조건 학습을 적용한다. 본 개시에 따르면, 베이스라인 다중 객체 추적 방법보다 더 나은 성능을 얻을 수 있다.
- [0013] 도 1은 본 개시에 따른 컴퓨터 장치(100)를 개략적으로 도시하는 도면이다. 도 2는 본 개시에 따른 컴퓨터 장치(100)에 의한 객체 전체 모델들 및 모델 상태 기반 온라인 제한 학습을 통한 효율적인 다중 객체 추적의 전체 프레임워크를 도시하는 도면이다.
- [0014] 도 1을 참조하면, 컴퓨터 장치(100)는 효율적인 다중 객체 추적을 위해 제공되며, 입력 모듈(110), 출력 모듈(120), 메모리(130), 또는 프로세서(140) 중 적어도 하나를 포함할 수 있다. 어떤 실시예에서, 컴퓨터 장치(100)의 구성 요소들 중 적어도 하나가 생략될 수 있으며, 적어도 하나의 다른 구성 요소가 추가될 수 있다. 어떤 실시예에서, 컴퓨터 장치(100)의 구성 요소들 중 적어도 두 개가 하나의 통합된 회로로 구현될 수 있다.
- [0015] 입력 모듈(110)은 컴퓨터 장치(100)의 적어도 하나의 구성 요소에 사용될 신호를 입력할 수 있다. 입력 모듈(110)은, 사용자가 컴퓨터 장치(100)에 직접적으로 신호를 입력하도록 구성되는 입력 유닛, 주변의 변화를 감지하여 신호를 발생하도록 구성되는 센서 유닛, 또는 외부 장치로부터 신호를 수신하도록 구성되는 수신 유닛 중 적어도 하나를 포함할 수 있다. 예를 들면, 입력 유닛은 마이크로폰(microphone), 마우스(mouse) 또는 키보드(keyboard) 중 적어도 하나를 포함할 수 있다. 어떤 실시예에서, 입력 유닛은 터치를 감지하도록 설정된 터치 회로(touch circuitry) 또는 터치에 의해 발생하는 힘의 세기를 측정하도록 설정된 센서 회로 중 적어도 하나를 포함할 수 있다.
- [0016] 출력 모듈(120)은 컴퓨터 장치(100)의 외부로 정보를 출력할 수 있다. 출력 모듈(120)은, 정보를 시각적으로 출력하도록 구성되는 표시 유닛, 정보를 오디오 신호로 출력할 수 있는 오디오 출력 유닛, 또는 정보를 무선으로 송신할 수 있는 송신 유닛 중 적어도 하나를 포함할 수 있다. 예를 들면, 표시 유닛은 디스플레이, 홀로그램 장치 또는 프로젝터 중 적어도 하나를 포함할 수 있다. 일 예로, 표시 유닛은 입력 모듈(110)의 터치 회로 또는 센서 회로 중 적어도 하나와 조립되어, 터치 스크린으로 구현될 수 있다. 예를 들면, 오디오 출력 유닛은 스피커 또는 리시버 중 적어도 하나를 포함할 수 있다.
- [0017] 일 실시예에 따르면, 수신 유닛과 송신 유닛은 통신 모듈로 구현될 수 있다. 통신 모듈은 컴퓨터 장치(100)에서 외부 장치와 통신을 수행할 수 있다. 통신 모듈은 컴퓨터 장치(100)와 외부 장치 간 통신 채널을 수립하고, 통신 채널을 통해, 외부 장치와 통신을 수행할 수 있다. 여기서, 외부 장치는 위성, 기지국, 또는 서버 중 적어도 하나를 포함할 수 있다. 통신 모듈은 유선 통신 모듈 또는 무선 통신 모듈 중 적어도 하나를 포함할 수 있다. 유선 통신 모듈은 외부 장치와 유선으로 연결되어, 유선으로 통신할 수 있다. 무선 통신 모듈은 근거리 통신 모듈 또는 원거리 통신 모듈 중 적어도 하나를 포함할 수 있다. 근거리 통신 모듈은 외부 장치와 근거리 통신 방식으로 통신할 수 있다. 예를 들면, 근거리 통신 방식은, 블루투스(Bluetooth), 와이파이 다이렉트(WiFi

direct), 또는 적외선 통신(IrDA; infrared data association) 중 적어도 하나를 포함할 수 있다. 원거리 통신 모듈은 외부 장치와 원거리 통신 방식으로 통신할 수 있다. 여기서, 원거리 통신 모듈은 네트워크를 통해 외부 장치와 통신할 수 있다. 예를 들면, 네트워크는 셀룰러 네트워크, 인터넷, 또는 LAN(local area network)이나 WAN(wide area network)과 같은 컴퓨터 네트워크 중 적어도 하나를 포함할 수 있다.

[0018] 메모리(130)는 컴퓨터 장치(100)의 적어도 하나의 구성 요소에 의해 사용되는 다양한 데이터를 저장할 수 있다. 예를 들면, 메모리(130)는 휘발성 메모리 또는 비휘발성 메모리 중 적어도 하나를 포함할 수 있다. 데이터는 적어도 하나의 프로그램 및 이와 관련된 입력 데이터 또는 출력 데이터를 포함할 수 있다. 프로그램은 메모리(130)에 적어도 하나의 명령을 포함하는 소프트웨어로서 저장될 수 있으며, 운영 체제, 미들 웨어 또는 어플리케이션 중 적어도 하나를 포함할 수 있다.

[0019] 프로세서(140)는 메모리(130)의 프로그램을 실행하여, 컴퓨터 장치(100)의 적어도 하나의 구성 요소를 제어할 수 있다. 이를 통해, 프로세서(140)는 데이터 처리 또는 연산을 수행할 수 있다. 이 때, 프로세서(140)는 메모리(130)에 저장된 명령을 실행할 수 있다. 다양한 실시예들에 따르면, 프로세서(140)는 도 2에 도시된 바와 같이 객체 전체 모델들 및 모델 상태 기반 온라인 제한 학습을 통해 효율적인 다중 객체 추적을 수행할 수 있다.

[0020] 프로세서(140)는 효율적인 다중 객체 추적을 위해, 객체 전체 모델들을 사용할 수 있다. 본 개시는 효율적인 다중 객체 추적을 위해 낮은 학습 및 추론 복잡성을 유지하면서 추적되는 서로 다른 객체들을 정확하게 구별할 수 있는 객체 전체 모델들, 바꿔 말해, 글로벌 객체 모델들을 제안한다. 구체적으로, 프로세서(140)는 온라인 다중 객체 추적을 수행하면서 연속 프레임들 사이에서 감지들을 연계시키기 위해 객체 전체 모델들을 사용할 수 있다. 이 때, 객체 전체 모델들은 객체 전체 모델들은 글로벌 외관 모델(global appearance model) 및 글로벌 상대 모션 모델(global relation motion model)을 포함할 수 있다. 이에 대해, 보다 상세하게 후술될 것이다.

[0021] 먼저 대조 학습을 사용하여 글로벌 외관 모델이 설계된다. 구체적으로, 가벼운 ConvNet에서 추적된 객체들에 대한 높은 수준의 시맨틱 객체 특징들을 추출한다. 그런 다음, 식별들을 고려하여 포지티브(positive)(즉, 샘플) 및 네거티브(negative)(즉, 다른) 객체 특징 쌍들을 정의한다. 트리플렛 손실(triplet loss)을 기반으로, 포지티브 객체 쌍들 사이의 특징 거리를 최소화하고 네거티브 객체 쌍들의 특징 거리를 최대화한다. 또한, 이전 모션들에서 객체의 미래 궤적을 예측하는 글로벌 상대 모션 모델을 제시한다. 글로벌 상대 모션 모델의 주요 모션 큐들로서 각 객체의 과거 궤적 및 추적된 객체들 간의 상대 모션을 사용합니다. 모션 예측 결과들을 개선하기 위해 글로벌 상대 모션 모델은 주로 생성기와 판별기로 구성되고 적대적 학습을 사용하여 그들을 훈련한다. 글로벌 외관 및 모션 모델들이 효과적인 객체 모델들이지만, 각 프레임에서 이러한 모델들을 온라인으로 학습하면 다중 객체 추적의 복잡성이 증가한다.

[0022] 따라서, 프로세서(140)는 이러한 모델들을 적응적으로 업데이트하는 객체 제한 학습을 적용할 수 있다. 주요 아이디어는 해당 모델들의 식별 가능성이 추적된 객체들을 구별하기에 불충분할 때 이러한 모델들을 업데이트하는 것이다. 이를 달성하기 위해, 모델 식별 가능성의 객체 제약 조건을 정의하고 제약 조건과 불일치가 발생하면 객체 모델들을 업데이트한다. 그 대가로, 온라인 학습 복잡성도 줄일 수 있다. 신뢰도 기반 데이터 연계를 위해 글로벌 외관 및 모션 모델들과 객체 제약 조건 학습을 적용한다. 본 개시에 따르면, 베이스라인 다중 객체 추적 방법보다 더 나은 성능을 얻을 수 있다.

[0023] 온라인 다중 객체 추적

[0024] 이 섹션에서는 온라인 다중 객체 추적 방법에 대해 설명한다. 친화도(affinity) 모델들과의 데이터 연계는 고품질 다중 객체 추적을 달성하기 위해 중요하다. 이를 위해 신뢰도 기반 객체 연계 알고리즘과 글로벌 모델들과의 친화도 모델들을 활용한다.

[0025] 신뢰도 기반 객체 연계(도 2의 (a) 참조)

[0026] 이 섹션에서는 기본 데이터 연계 방법으로 신뢰도 기반 다중 객체 추적을 소개한다. 프레임 t 에서 객체 디텍터의 감지들을 $z_t = [z_x, z_y, z_w, z_h]$ 로 나타낸다. 여기서, z_x, z_y, z_w, z_h 는 각각 상자(box)의 x 및 y 위치, 폭, 높이다. 감지 집합(set of detection)을 프레임 t 에서 Z_t 로 정의한다. O^i 를 트랙렛(tracklet)으로 나타내며, 그

것은 각 프레임에서 z_t 와 연계될 수 있다. 따라서, 트랙렛

$$O^i = \{z_k^i | 1 < t_s^i < k < t_e^i \leq t\}$$

는 시작과 종료 타임스탬프를 나타내는 t_s^i 와 t_e^i 사이의 짧은 궤적이다. 따라서, 프레임별 온라인 연계의 주요

문제는 O^i 를 이 트랙렛에서 유래한 $\mathbf{z}_t^i$ 와 연계하는 방법이다.

[0027] 신뢰도 연계의 경우 길이와 연속성으로 각 O^i 의 신뢰도를 평가하고 연계된 감지와의 친화도를 하기 [수학식 1]과 같이 평가할 수 있다.

수학식 1

$$C(O^i) = \left(\frac{1}{L} \sum_{k \in [t_s^i, t_e^i], v^i(k)=1} A(O^i, \mathbf{z}_k^i) \right) \times \left(1 - \exp \left(-\beta \cdot \sqrt{(L - \lambda)} \right) \right),$$

[0028]

[0029] 여기서, C는 트랙렛 O^i 에 대한 신뢰도 함수이고 $v^i(t)$ 는 O^i 와 $\mathbf{z}_k^i$ 사이의 연계 사건을 나타내는 이진 함수이다. 따라서, $v^i(t) = 1$ 은 객체 i에 대한 연계된 감지 $\mathbf{z}_k^i$ 가 프레임 t에 존재함을 의미하며, 그렇지 않으면 $v^i(t) = 0$ 이다. $A(O^i, \mathbf{z}_k^i)$ 는 하기 [수학식 4]로 계산되는 총 친화도 점수이다. L은 $L = IO^i$ 로서 트랙렛의 길이이고, β 는 디텍터의 정확도에 의존하는 제어 파라미터이다. $\lambda = t_e^i - t_s^i + 1 - L$ 은 다른 객체들에 의한 폐색 또는 신뢰할 수 없는 감지로 인해 객체 i가 누락된 스킵 프레임들의 수이다.

[0030] 트랙렛의 신뢰도 점수가 상기 [수학식 1]에 의해 계산될 때, 그것의 신뢰도에 따라 로컬 및 글로벌 연계를 적응적으로 수행한다. 신뢰도가 높은 트랙렛 $O^{i(high)}$ 는 신뢰성이 높은 트랙렛으로 볼 수 있지만 신뢰도가 낮은 트랙렛은 신뢰도가 낮은 트랙렛으로 간주된다. $O^{i(high)}$ 는 감지와 로컬하게 연계되어 있다. 프레임 t에서 높은 신뢰도를 갖는 h 개의 트랙렛들과 n개의 감지들 $\mathbf{Z}_t = \{\mathbf{z}_t^j\}_{j=1}^n$ 이 주어지면 하기 [수학식 2]와 같이 로컬 연계 행렬 S_{local} 을 계산한다.

수학식 2

$$\mathbf{S}_{h \times n}^{local} = [s_{ij}], s_{ij} = -A(O^{i(high)}, \mathbf{z}_t^j), \mathbf{z}_t^j \in \mathbb{Z}_t$$

[0031]

[0032] 논의된 바와 같이, 낮은 신뢰도의 $O^{i(low)}$ 를 갖는 트랙렛은 신뢰할 수 없거나 단편화된 트랙렛으로 간주된다. 따라서, 이 단편화된 트랙렛을 다른 $O^{i(high)}$ 또는 감지 $\mathbf{z}_t^j$ 와 연결한다. 여기서, $\mathbf{z}_t^j$ 는 로컬 연계의 $O^{i(high)}$ 와 연계되지 않아야 한다. η 개의 비연계 감지들 ($\eta \leq n$)과 높은 신뢰도 및 낮은 신뢰도를 각각 갖는 h 개 및 l 개의 트랙렛들을 가정한다. $O^{i(low)}$ 를 연계하기 위해, 다음의 세 가지 가능한 이벤트들을 고려하여 글로벌 연계를 수행한다. 먼저, $O^{i(low)}$ 가 $O^{i(high)}$ 와 연계될 때, 이벤트 A를 나타낸다. $O^{i(low)}$ 가 종료되면 이벤트 B를 나타낸다. 마지막으로, $O^{i(low)}$ 가 $\mathbf{z}_t^j$ 와 연계될 때 이벤트 C를 나타낸다.

[0033] 그런 다음, 이러한 연계 이벤트들을 기반으로 글로벌 연계 행렬을 하기 [수학식 3]과 같이 정의한다.

수학식 3

$$\mathbf{S}_{(l+\eta) \times (h+l)}^{global} = \begin{pmatrix} A_{l \times h} & B_{l \times l} \\ -\theta_{\eta \times h} & C_{\eta \times l} \end{pmatrix}$$

여기서, $A = [a_{ij}]$ 는 이벤트 A 에 해당하고 $a_{ij} = -A(O^{i(low)}, O^{j(high)})$ 는 친화도 모델(하기 [수학식 4])로 평가된 연계 점수이다. 이벤트 B 는 $B = \text{diag}[b_1, \dots, b_l]$ 로 모델링되며, 여기서, $b_i = 1 - C(O^{i(low)})$ 는 $O^{i(low)}$ 가 끝나는 점수이다. 이벤트 C 는 $c_{ij} = -A(O^{i(low)}, \mathbf{z}_t^j)$ 에 의해 결정된다.

S^{local} 또는 S^{global} 이 있다면, 헝가리안(Hungarian) 알고리즘을 사용하여 각 행렬에서 최적의 일치 쌍을 결정할 수 있다.

친화도 모델(도 2의 (b) 참조)

트랙 신뢰도(상기 [수학식 1])과 신뢰도 기반 연계(상기 [수학식 2] 및 [수학식 3])의 평가를 위해서는 객체들 간의 친화도들을 평가하는 것이 중요하다. 친화도들을 정확하게 계산하기 위해, 몇 가지 객체 모델들을 활용한 다. 트랙렛 O^i 의 객체 모델들을 $\{A, S, M\}$ 으로 정의한다. 여기서, A, S, M 은 각각 외관(appearance), 형상(shape), 모션(motion) 모델들이다. 이러한 모델들을 사용하여, 전체 친화도 모델을 하기 [수학식 4]와 같이 정의한다.

수학식 4

$$\mathcal{A}(u, v) = \mathcal{A}^A(u, v) \cdot \mathcal{A}^S(u, v) \cdot \mathcal{A}^M(u, v)$$

여기서, u 와 v 는 트랙렛 또는 감지이다. 외관 친화도는 하기 [수학식 5]와 같이 정의된다.

수학식 5

$$\mathcal{A}^A(u, v) = \exp\left(-\frac{1}{c_A} \cdot d(\mathbf{f}^u, \mathbf{f}^v)\right)$$

여기서, c_A (예: 47.5)는 이 친화도를 조정하기 위한 하이퍼 파라미터이며, $d(\cdot)$ 은 특징 거리를 평가하기 위한 L2 노름(norm)이다. $\mathbf{f}^u$ 와 $\mathbf{f}^v$ 는 각각 u 와 v 의 외관 특징들이다. 이는 후술되는 바와 같이 글로벌 외관 모델에서 추출될 수 있다.

모션들로 객체들 간의 친화도 점수를 계산하기 위해, 하기 [수학식 6]과 같이 자체 및 상대 모션들을 기반으로 하는 새로운 모션 친화도 $\mathcal{A}^M(u, v)$ 을 제안한다.

수학식 6

$$\mathcal{A}^M(u, v) = c_M \mathcal{A}_{relation}^M(u, v) + (1 - c_M) \mathcal{A}_{self}^M(u, v)$$

여기서, $\mathcal{A}_{self}^M(u, v)$ 는 u 와 v 의 자체 모션들을 갖는 모션 친화도이다. 먼저 칼만(Kalman) 필터를 사용하여 객체의 자체 모션을 예측하고, 모션의 차이를 하기 [수학식 7]과 같이 평가한다.

수학식 7

$$\begin{aligned} \mathcal{A}_{self}^M(u, v) = & \mathcal{N}(\hat{\mathbf{b}}_u^{tail} + \mathbf{v}_u^F \Theta; \hat{\mathbf{b}}_v^{head} \Sigma_{self}^F) \\ & \times \mathcal{N}(\hat{\mathbf{b}}_v^{tail} + \mathbf{v}_v^B \Theta; \hat{\mathbf{b}}_u^{head} \Sigma_{self}^B) \end{aligned}$$

여기서, $\hat{\mathbf{b}}$ 는 칼만 필터에 의해 업데이트된 상태들이다. 이를 계산하기 위해, 시간 간격 Θ 로 u 와 v 헤드의 머리부터 꼬리까지의 공간적 차이를 사용한다. $\mathbf{v}_u^F$ 는 u 의 머리부터 꼬리까지 계산되는 전방 속도이다. 후방 속도 $\mathbf{v}_v^B$ 는 v 의 꼬리부터 머리까지 계산된다. 속도와 업데이트된 위치들 사이의 공간 거리 친화도를 평가하기 위해 가우스 분포 함수를 사용한다.

또한, 글로벌 상대 모션 모델을 기반으로 $\mathcal{A}_{relation}^M(u, v)$ 를 또 다른 모션 친화도 모델로 사용한다. 일부 객체 추적(예: 보행자, 자동차 등)에서는 상호 작용 및 그룹 행동으로 인해 발생하는 상대 모션을 활용하는 것이 유용하다. 객체 간의 이러한 상대 모션은 자체 모션 모델에 의해 포착되지 않지만 다중 객체 추적에 중요하기 때문에 친화도 평가를 위해 상대 모션을 활용하고 GAN을 사용하여 글로벌 상대 모션 모델을 학습한다. 이 관계 모델을 사용하여 각 객체 모션을 예측하는 세부 사항은 후술될 것이다.

글로벌 관계 모델의 주요 이점은 특정 프레임에서 다음 Δ_{est} 프레임까지의 다른 객체 모션들을 고려하여 추적된 모든 객체들의 모션들을 예측할 수 있다는 것이다. 따라서, 프레임당 이 모델을 학습할 필요가 없으며, 모션 추론 복잡성을 감소시킨다. 자체 및 상대 모션 모델들을 효과적으로 연계하기 위해, 가중치 파라미터 c_M 을 도입한다. 관계 모델의 모션 예측 $1 \leq \Delta < \Delta_{est}$ 범위를 고려하여 c_M 을 계산한다. 글로벌 상대 모션 모델이 향후 모션을 추정할 때, Δ 를 1로 설정한다. 글로벌 관계 모델이 새로운 미래 모션을 다시 예측할 때까지 Δ 는 증가한다. c_M 은 다음과 같이 계산한다.

수학식 8

$$c_M = \frac{(\Delta_{est} - \Delta)}{\Delta_{est}}$$

Δ 가 낮을 때, 자체 모션 모델보다 관계 모델에 더 높은 가중치를 할당한다. 그 이유는 Δ 가 증가할수록 추정된 모션의 정확도가 감소하는 경향이 있기 때문이다.

[0052] 이 친화도 모델 $\mathcal{A}_{relation}^M(u, v)$ 는 하기 [수학식 9]와 같이 정의될 수 있다.

수학식 9

$$\mathcal{A}_{relation}^M(u, v) = \mathcal{N}(\hat{Y}_u^\Theta; \hat{\mathbf{b}}_v^{head}, \Sigma_{relation}^F)$$

[0054] 여기서, N 은 가우스 분포이고, $\hat{Y}_u^\Theta$ 는 모션 관계를 고려한 글로벌 상대 모션 모델에서 업데이트된 상태들이다. $\hat{\mathbf{b}}_v^{head}$ 는 v 의 첫 번째 개선된 위치이고, $\Sigma_{relation}^F$ 는 커나이번스(connivance) 행렬이다. 여기서는, 추적 속도를 높이기 위해 전진 모션을 활용한다.

[0055] 객체 크기의 불일치를 고려하기 위해, 친화도 평가를 위한 형상 모델을 사용한다. 형상 친화도는 하기 [수학식 10]과 같이 정의된다.

수학식 10

$$\mathcal{A}^S(u, v) = \exp\left(-c_s \cdot \left\{ \frac{h_u - h_v}{h_u + h_v} + \frac{w_u - w_v}{w_u + w_v} \right\}\right)$$

[0057] 여기서, h 와 w 는 u 와 v 의 높이와 폭을 나타내며, c_s (예: 1.5)는 형상 친화도에 대한 조정 파라미터이다.

[0058] **글로벌 외관 모델**(도 2의 (d) 참조)

[0059] 견고한 연계를 달성하기 위해 객체 외관은 중요한 큐이다. 특히 효과적인 외관 모델을 활용하면 폐색 및 외관 변화 시 연계 실패를 줄일 수 있다. 이 섹션에서는 글로벌 외관 모델에 대해 설명한다. 경량 ConvNet을 사용하여 추적된 객체들에 대한 높은 수준의 의미론적 객체 특징들을 추출한다. 추적된 객체들의 식별들을 통해 포지티브(positive) 객체 및 네거티브(negative) 객체 특징 쌍들을 정의한다. 또한, 포지티브 쌍들 사이의 특징 거리를 최소화하고, 네거티브 쌍들 사이의 특징 거리를 최대화하기 위해 트리플릿(triplet) 손실을 이용한다.

[0060] 심층 특징 추출기

[0061] 도 3은 본 개시에 따른 글로벌 외관 모델의 심층 특징 추출기의 아키텍처를 도시하는 도면이다.

[0062] 도 3을 참조하면, 객체의 외관 특징을 추출하기 위해 LuNet이라고 하는 수정된 ResNet-v2 네트워크를 사용한다. LuNet의 입력은 128×64 이미지 패치이다. 이 네트워크는 LeakyReLU를 활성화 함수로 사용한다. 스트라이드 컨볼루션 대신 스트라이드 2를 갖는 다중 3×3 최대 풀링이며, 도 3에 도시된 바와 같이 마지막 res-블록에서 피쳐 맵의 최종 평균 풀링 계층을 제거한다. 마지막 계층의 MLP 계층에서, 객체의 128차원 임베딩 특징을 추출한다. 이 네트워크는 다른 특징 추출 네트워크(예: ResNet-50)에 비해 경량(5M 파라미터)이다. 식별 가능한 특징들을 학습하기 위해 이 네트워크를 트리플릿 손실로 훈련하고, 훈련 방법에 대한 세부 정보는 후술될 것이다.

[0063] 트리플릿 손실

[0064] 연속 프레임들에서 추적된 객체들을 식별하기 위해서는 객체 외관 특징들 사이의 거리를 정확하게 계산하는 것이 중요하다. 이를 위해, 트리플릿 손실을 기반으로 글로벌 외관 모델을 훈련한다. 도 4는 본 개시에 따른 글로벌 외관 모델에서 트리플릿 손실을 사용하여 객체 특징 거리들을 변경하는 예를 도시하는 도면이다. 앵커(anchor) x_a^i , 포지티브 x_p^i 및 네거티브 x_n^i 객체들을 ID들로 정의한다. 앵커 객체는 타겟 객체를 의미한다. 포지티브 객체는 앵커 객체와 동일한 ID를 가지지만 네거티브 객체는 다른 ID를 갖는다. 특징 추출기 $f(\cdot)$ 를 사용하여 128차원 임베딩 특징들을 추출할 수 있으며, $f(x_a^i)$, $f(x_p^i)$, 및 $f(x_n^i)$ 를 앵커, 포지티브, 및 네거티브

브 특징들로 나타낸다. 객체들 간의 차별성을 높이기 위해서는 도 4에 도시된 바와 같이 x_a^i 와 x_p^i 사이의 거리를 최소화하고 x_a^i 와 x_n^i 사이의 거리를 최대화해야 한다. 이를 위해 트리플릿 손실을 하기 [수학식 11]과 같이 정의할 수 있다.

수학식 11

$$\mathcal{L}_{triplet} = \max\left(0, m + d\left(f(x_a^i), f(x_p^i)\right) - d\left(f(x_a^i), f(x_n^i)\right)\right)$$

여기서, $d(\cdot)$ 는 두 임베딩 벡터들 사이의 거리 함수이고, m 은 포지티브 샘플과 네거티브 샘플 사이의 거리를 강제하는 마진이다.

온라인 하드(hard) 트리플릿 마이닝(mining) 및 손실

기본적으로 트리플릿 손실은 앵커/포지티브 및 앵커/네거티브 쌍들의 거리들로 평가된다. 따라서, 샘플 쌍 선택은 이 메트릭 학습에 크게 영향을 미친다. 온라인 추적 중 의미 있는 샘플 쌍들을 결정하기 위해, 온라인 트리플릿 마이닝을 사용한다. 온라인 트리플릿 마이닝에서, 각 배치(batch)에서 가능한 트리플릿들을 선택한다. 특징 추출기에서 추적된 객체들에 대해 임베딩 특징 벡터들이 추출되면, 객체 ID들을 사용하여 샘플 트리플릿들을 구성한다. 서로 다른 샘플 조합들을 기반으로, 모든 샘플 특징들을 앵커, 포지티브, 네거티브로 사용할 수 있다. 배치에서 훈련 샘플 트리플릿들을 일괄적으로 무작위로 선택하고 메트릭 학습(상기 [수학식 11])에 사용할 수 있다. 그러나, 보다 효과적인 학습을 위해 온라인 하드 트리플릿 마이닝을 제시한다. 기본 아이디어는 앵커를 기준으로 가장 하드한(hardest) 포지티브 및 네거티브 샘플을 사용하는 것이다. 여기서, 하드 포지티브 샘플 배치 내의 포지티브 샘플들 중 가장 낮은 친화도(상기 [수학식 4])를 갖는다. 하드 네거티브 샘플은 네거티브 샘플들 중에서 가장 높은 친화도를 갖는다. 더 나쁜 경우, 포지티브 쌍의 친화도는 네거티브 쌍의 친화도보다 낮을 가능성이 있다. 이러한 하드 트리플릿들은 무작위 훈련 샘플 선택보다 수렴 시간을 줄이고 모델 정확도를 높이는 데 더 효과적이다.

이 아이디어를 바탕으로, 기존의 트리플릿 손실(상기 [수학식 3])을 하기 [수학식 12]와 같이 하드 트리플릿 손실로 변환할 수 있다.

수학식 12

$$\mathcal{L}_{hard} = \sum_{i=1}^P \sum_{a=1}^K \left[m + \max_{p=1 \dots K} d\left(f(x_a^i), f(x_p^i)\right) - \min_{\substack{j=1 \dots P \\ n=1 \dots K \\ j \neq i}} d\left(f(x_a^i), f(x_n^j)\right) \right]_+$$

여기서, $[\cdot]_+$ 는 힌지 손실이고, P 는 구별되는 객체 ID들의 수이며, K 는 클래스당 샘플 이미지들의 수를 의미한다. i 와 j 는 클래스 인덱스들($i \neq j$)이다. 두 번째 항과 세 번째 항은 각각 가장 하드한 포지티브 샘플과 가장 하드한 네거티브 샘플 선택을 나타낸다. 모델 $f(\cdot)$ 를 훈련하는 데 이러한 어려운 샘플들을 사용함으로써 모델의 판별성을 더욱 향상시킬 수 있다. 하드 포지티브 및 네거티브 샘플들에 대한 몇 가지 예가 도 5에 제공한다.

글로벌 상대 모션 모델(도 2의 (e) 참조)

다중 객체 추적의 높은 품질을 달성하기 위해서는 객체 미래 궤적을 정확하게 예측하는 것이 중요하다. 전술된 바와 같이, 모션 친화도를 평가하기 위해 자체 및 상대 모션 모델들을 사용한다. 이 섹션에서는 상대 모션 모델

과 기울기 방법을 자세히 제시한다.

[0074] 상대 모션을 고려하는 것은 미래 궤적들을 더 정확하게 추정하기 위한 핵심 아이디어 중 하나이다. 객체(예: 보행자들, 장애물들 등)의 움직임은 추적하는 동안 인근 객체들과 장애물들에 의해 영향을 받을 수 있기 때문이다. 이러한 이유로, 일부 최근 연구는 LSTM과 같은 비선형 모델과의 상대 모션을 고려하여 다중 객체의 궤적을 추정한다. 따라서, 추적된 모든 객체들에 대한 글로벌 모션들을 추정하기 위한 상대 모션을 사용한다.

[0075] 미래 궤적을 추정하기 위해, 글로벌 상대 모션 모델은 연속적인 δ_{obs} 개의 프레임들을 활용한다(예: $\delta_{obs} = 5$). 따라서, 이 모델의 입력은 이전 $t - \delta_{obs} - 1$ 에서 현재 t 까지의 프레임들의 모션 궤적들 $\mathbf{X} = \{X_{t-\delta_{obs}-1}, \dots, X_t\}$ 의 집합이며, 출력은 다음 $t + \Delta_{est}$ 의 프레임들에 대한 추정 궤적들 $\hat{\mathbf{Y}} = \{\hat{Y}_{t+1}, \dots, \hat{Y}_{t+\Delta_{est}}\}$ 의 집합이다. X_t 와 $\hat{Y}_T$ 는 각각 이전 프레임과 미래 프레임에 대한 x 와 y 좌표들로 구성된 추적된 객체 궤적들이다.

[0076] 글로벌 상대 모션 모델은 다양한 상대 모션의 분포를 이해하기 위해 적대적 학습을 기반으로 훈련된다. 적대적 학습에 대한 세부 사항은 후술될 것이다. 본 개시의 모델은 궤적 생성기 G 와 판별기 D 의 두 개의 네트워크로 구성된다. 생성기 G 는 모션 이력의 분포를 포착하고 미래의 모션들을 추정한다. 판별기 D 는 실측 궤적 Y 또는 예측 궤적 $\hat{Y}$ 에 대한 신뢰도 점수를 계산한다. G 와 D 의 상세한 구조는 후술될 것이다.

[0077] 적대적 생성망들

[0078] 이 섹션에서는 모션 학습 방법을 소개하기 전에 먼저 전통적인 적대적 생성망을 간략하게 검토한다. 기본적으로 생성기와 판별기는 다른 네트워크의 피드백을 교환하여 훈련된다. 따라서, 생성기 G 는 훈련된 데이터셋의 샘플 분포를 포착하고 판별기 D 를 속이기 위한 보다 현실적인 샘플들을 생성하려고 한다. 이상적으로, 생성된 샘플의 분포는 실제 샘플의 분포에 가깝다. 생성기 G 는 잠재 벡터 z 를 입력으로 취하고 생성된 샘플 $G(z)$ 를 출력한다. 반면에, 판별기는 훈련된 집합의 실제 또는 생성기에서 생성된 가짜 샘플을 취한다. $D(\cdot)$ 의 출력은 샘플의 신뢰도 점수이다. 따라서, 이러한 적대적 학습의 손실은 일반적으로 2인용 미니맥스 게임을 기반으로 하기 [수학식 13]과 같이 표현된다.

수학식 13

$$\mathcal{L}_{adv} = \min_G \max_D \mathbb{E}_{\mathbf{x} \sim p_{data}(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log (1 - D(G(\mathbf{z})))]$$

[0079]

[0080] 여기서, p_{data} 와 p_z 는 각각 주어진 데이터셋과 생성된 샘플의 분포들이다. 글로벌 모션 학습을 위해 이 손실을 연장할 수 있다. 궤적 생성을 위해, 하기 [수학식 14]와 같이 적대적 손실 $\mathcal{L}_M$ 을 정의한다.

수학식 14

$$\mathcal{L}_M = \min_G \max_D \mathbb{E}_{\mathbf{Y} \sim p_{data}(\mathbf{Y})} [\log D(\mathbf{Y})] + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log (1 - D(G(\mathbf{z}, \mathbf{X}^*)))]$$

[0081]

[0082] 여기서, Y 는 실측 모션이며, X^* 는 Y 보다 높은 정밀도로 선택된 궤적이다.

[0083] 글로벌 상대 모션 모델을 훈련시키는 주요 어려움 중 하나는 제한된 샘플들(즉, 궤적들)로 인한 불안정한 모션

추정이다. 사실, 객체 모션들은 너무 다양하지만, 추적하는 동안 수집된 궤적은 모든 사례들을 처리하기에 매우 제한적이다. 모션 모델들이 다양한 모션들을 예측하도록 장려하기 위해, N_T 개의 시간들 동안 $N(0, 1)$ 을 갖는 잠재 벡터 z 를 무작위로 샘플링한다(예: $N_T = 20$). 따라서, 생성기에서 하나의 샘플당 N_T 개의 궤적들을 생성할 수 있다. 생성된 모든 샘플들을 훈련에 사용할 수 있지만, 민감도 문제를 처리하기 위해 Top-1 샘플링을 사용한다. 실제 궤적에 가장 가까운 것 X^* 를 선택하고, 선택된 것은 적대적 학습(상기 [수학식 14])을 위한 것이다. Top-1 샘플링을 사용하면 생성된 샘플들을 모두 사용하는 것보다 예측된 모션들의 다양성을 높일 수 있다.

[0084] 생성기 및 판별기

[0085] 이 섹션에서는 생성기 G 와 판별기 D 의 구조들에 대해 논의한다. 도 6은 본 개시에 따른 글로벌 상대 모션 모델의 아키텍처를 도시하는 도면이다. 도 6을 참조하면, 글로벌 상대 모션 모델이 생성기 G 와 판별기 D 를 갖는다. 각 네트워크에 대한 자세한 내용은 다음과 같다.

[0086] <생성기>

[0087] 온라인 추적 중 각 객체의 모션들을 추정하기 위해 생성기 G 가 활용된다. 다중 객체 추적 중 추적된 모든 객체들의 x 와 y 좌표들로 X 를 구성한다. 생성기 G 는 인코더, 풀링 모듈, 디코더의 세 개의 네트워크들로 구성된다. 각 네트워크에 대한 자세한 내용은 다음과 같다.

[0088] **인코더:** 인코더로부터, 이전의 모션들과 히든(hidden) 상태들로 현재 프레임 t 에서 각 객체 i 의 히든 상태들

$h_t^{i(e)}$ 을 학습할 수 있다. 인코더는 LeakyReLU를 갖는 완전 연결 계층으로 모션 임베딩 벡터를 추출하고, LSTM 을 사용하여 객체 i 의 히든 상태들 $h_t^{i(e)}$ 을 출력한다. 그런 다음, 학습된 히든 상태들 $h_t^{i(e)}$ 을 풀링 네트워크에 공급한다.

[0089] **풀링 모듈:** 객체들 간의 상대 모션들을 고려하기 위해 풀링 모듈을 사용한다. 풀링 모듈은 먼저 객체와 다른 객체들 사이의 상대 위치들을 계산한다. 이러한 상대 위치들은 타겟 객체와 다른 객체들의 x 와 y 좌표들을 빼서 계산하며, 각 객체의 히든 상태들과 연결된다. 그런 다음, 그것들은 LeakyReLU를 갖는 MLP에 의해 독립적으로 임베딩된다. 마지막으로, 임베딩 벡터들은 풀링(예: 최대 풀링을 사용함)되어 각 객체 i 의 풀링 벡터 P_t^i 가 계산된다. 풀링 벡터 P_t^i 을 사용하여, 상대 정보를 요약하고, 디코더에서 미래 궤적을 예측하는 데 사용할 수 있다.

[0090] **디코더:** 디코더는 두 개의 MLP들과 하나의 LSTM 네트워크로 구성된다. 이것의 입력으로, 인코더와 풀링 모듈의

출력을 사용한다. 히든 상태 $h_t^{i(e)}$, 풀링 벡터 P_t^i , 및 잠재 벡터 z 을 연결한다. 이 특징은 객체 모션 $h_t^{i(e)}$, 및 객체와 다른 객체 P_t^i 사이의 상대 모션을 나타낸다. 그런 다음, 학습된 모션 분포들을 사용하여 객체의 방향과 속도를 고려하여 i 번째 객체 특징 모션을 추정하기 위해 z 이 추가로 임베딩된다. 연결된 특징들은 디코더의 LSTM으로 전달되고 이 네트워크는 히든 상태 $h_t^{i(d)}$ 를 출력한다. 마지막 MLP는 $h_t^{i(d)}$ 를 사용하여 각 객체의 미래 궤적들을 예측한다. 그런 다음, $h_t^{i(d)}$, P_t^i , 및 z 을 연결한다. 그런 다음, 디코더가 Δ_{est} 궤적들을 예측할 때까지 연결된 특징을 LSTM에 반복적으로 공급한다.

[0091] <판별기>

[0092] 판별기 D 는 분류기로서 인코더와 MLP로 구성된다. D 의 입력은 디코더에 의한 실측 궤적(즉, 실제)과 예측 궤적(즉, 가짜)이다. 그런 다음, D 의 인코더는 G 의 인코더와 유사한 입력 궤적에 대한 히든 특징들을 추출한다. 그런 다음, 히든 상태들을 재귀적으로 예측하여 더 세분화한다. 인코더의 마지막 히든 상태를 완전 연결 계층에 공급함으로써 입력 궤적이 실제인지의 여부를 나타내는 분류 신뢰도를 예측할 수 있다. D 의 신뢰도는 LM을 계산하는 데 사용된다(상기 [수학식 14]).

[0093] 객체 제한 학습(도 2의 (c) 참조)

[0094] 이 섹션에서는 글로벌 외관 모델의 업데이트 일정을 제어하기 위한 객체 제약 조건 학습 방법을 소개한다. 전술된 바와 같이, 모든 프레임들에서 글로벌 모델의 온라인 학습은 다중 객체 추적의 복잡성을 증가시킨다. 외관 특징들은 미래 프레임들에서 객체들 간의 친화도를 평가하기에 충분한 식별력을 가지고 있다고 가정한다. 이러한 가설에 따라 객체 모델들의 식별 가능성에 따라 적응적으로 모델들을 업데이트하여 추적 속도와 정확도를 동시에 향상시킬 수 있다. 이를 위해, 이 섹션에서는 객체 제약 학습을 제안한다. 이 아이디어는 간단하고 구현하기 쉽다. 제약 조건과 불일치가 발생할 때만 모델을 업데이트한다.

[0095] 글로벌 외관 모델의 업데이트들의 수를 줄이기 위해, 외관 특징들의 식별 가능성을 고려한다. 외관 특징의 식별 가능성을 계산하기 위해 연속 프레임들 간의 객체 외관 상관 관계를 사용한다. 객체들이 여러 프레임들 동안 급격한 외관 변화가 없으면 이 프레임들에서 추출된 외관 특징들은 유사하다. 따라서, 현재 프레임에서 추적하기 위해 과거 프레임의 외관 특징을 활용할 수 있다. 그러나, 외관의 변화나 폐색으로 인해 특징들의 유사도가 낮은 경우에는 현재 프레임에서 특징들을 업데이트해야 합니다. 이러한 상황들의 예가 도 7에 제공된다. 프레임 t 에서의 외관 모델의 식별 가능성을 측정하기 위해, 두 개의 서로 다른 트랙렛들 O_i 와 O_j 사이의 외관 친화도 $A^A(O^i, O^j)$ 로 ρ_t 를 계산한다. ρ_t 를 하기 [수학식 15]와 같이 정의한다.

수학식 15

$$\rho_t = 1 - \frac{1}{N_t^2 - N_t} \left(\sum_{i=1}^{N_t} \sum_{j \neq i}^{N_t} A^A(O^i, O^j) \right)$$

[0096]

[0097] 여기서, N_t 는 프레임 t 에서 트랙렛들의 수이다. 상기 [수학식 15]를 사용하여, 평균 외관 친화도를 계산한다. ρ_t 를 $[0, 1]$ 로 설정하기 위해, 각 $A^A(O^i, O^j)$ 에 대해 최소 최대 정규화를 수행한다.

[0098] ρ_t 가 1에 가까우면 객체들을 구별할 수 있는 충분한 식별력이 있다고 간주한다. 예를 들어, 임계값 μ 를 0.6으로 설정한다. $\rho_t > \mu$ 인 경우, 모델이 여전히 높은 식별성을 유지하기 때문에 외관 모델을 업데이트하지 않는다. 그렇지 않은 경우, 프레임 t 에서 모델을 업데이트한다. 또한, 새 객체가 나타나거나 기존 트랙이 종료되면 객체 외관들을 업데이트한다. 요약하면, 다음 두 가지 제약 조건들을 기반으로 글로벌 외관 모델을 업데이트한다: 제약 조건 1: ρ_t 가 μ 보다 낮을 것; 및 제약 조건 2: $N_t \neq N_{t-1}$ 일 것.

[0099] 도 8 및 도 9는 본 개시에 따른 알고리즘들을 나타내는 도면들이다. 구현의 용이함을 위해 글로벌 외관 모델을 사용하여 객체 제약 학습을 위한 도 8에 도시된 바와 같은 알고리즘 1을 제시한다. 또한, 도 9에 도시된 바와 같은 알고리즘 2에서 제안된 글로벌 모델과 객체 제약 학습을 기반으로 전반적인 다중 객체 추적 알고리즘을 제공한다.

[0100] 도 10은 본 개시에 따른 컴퓨터 장치(100)의 방법을 개략적으로 도시하는 도면이다.

[0101] 도 10을 참조하면, 컴퓨터 장치(100)는 210 단계에서 글로벌 외관 모델을 이용하여, 연속 프레임들에서 감지되는 객체들의 외관 특징들을 검출할 수 있다. 다양한 실시예들에 따르면, 글로벌 외관 모델은, 학습 시, 훈련 데이터셋에서 감지되는 객체들을 아이디(ID)들로 정의하고, 타겟인 앵커 객체, 앵커 객체와 동일한 아이디의 포지티브 객체, 및 앵커 객체와 다른 아이디의 네거티브 객체를 선택하고, 앵커 객체와 포지티브 객체의 외관 특징들에 따른 포지티브 객체 쌍들, 및 앵커 객체와 네거티브 객체의 외관 특징들에 따른 네거티브 객체 쌍들을 정의하고, 트리플렛 손실을 기반으로, 포지티브 객체 쌍들 사이의 특징 거리를 최소화하고, 네거티브 객체 쌍들 사이의 특징 거리를 최대화할 수 있다.

[0102] 다음으로, 컴퓨터 장치(100)는 220 단계에서 글로벌 상대 모션 모델을 이용하여, 감지되는 객체들의 미래 궤적들을 예측할 수 있다. 다양한 실시예들에 따르면, 글로벌 상대 모션 모델은, 훈련 데이터셋에서 감지되는 객체들의 각각의 과거 궤적 및 객체들 사이의 상대 모션을 기반으로 객체들의 각각의 미래 궤적을 예측하도록, 학습될 수 있다. 구체적으로, 글로벌 상대 모션 모델은, 과거 궤적 및 상대 모션을 기반으로 객체들의 각각의 미래

궤적을 생성하도록 구성되는 생성기, 및 예측된 미래 궤적이 훈련 데이터셋에서 확인되는 실측 궤적과 일치하는지의 여부를 판별하도록 구성되는 판별기를 포함할 수 있다.

[0103] 다양한 실시예들에 따르면, 글로벌 외관 모델 또는 글로벌 상대 모션 모델 중 적어도 하나는, 학습 시 미리 정해진 제약 조건을 기반으로 업데이트될 수 있다. 일 실시예에 따르면, 프로세서(140)는 새 객체가 나타나는 경우, 글로벌 외관 모델 및 글로벌 상대 모션 모델을 업데이트할 수 있다. 다른 실시예에 따르면, 프로세서(140)는 한 객체에 대해 상이한 프레임들에서 검출되는 외관 특징들의 유사도가 임계값 이하인 경우, 글로벌 외관 모델을 업데이트할 수 있다. 또 다른 실시예에 따르면, 프로세서(140)는 현재 프레임에서 예측되는 미래 궤적들의 수가 미리 정해진 수 미만이면, 글로벌 상대 모션 모델을 업데이트하도록 구성될 수 있다.

[0104] 다음으로, 컴퓨터 장치(100)는 230 단계에서 외관 특징들과 미래 궤적들을 기반으로 감지되는 객체들을 구별하여 연계시킬 수 있다. 이러한 방식으로, 프로세서(140)는 연속 프레임들에서 다중 객체를 추적할 수 있다.

[0105] 본 개시에 따르면, 객체 전체 모델들을 이용하여 효율적인 다중 객체 추적이 가능하다. 구체적으로, 글로벌 외관 모델에 의해 객체들에 대한 높은 식별성을 확보하고, 글로벌 상대 모션 모델에 의해 각 객체의 정확한 궤적 추정이 가능하므로, 다중 객체 추적의 추적 정확도 및 속도를 함께 향상시킬 수 있다. 또한, 객체 전체 모델들에 대한 객체 제한 학습을 통해, 온라인에서의 객체 전체 모델들을 학습하는 계산 비용이 완화될 수 있다.

[0107] 요컨대, 본 개시는 객체 전체 모델들 및 모델 상태 기반 온라인 제한 학습을 통한 효율적인 다중 객체 추적 방법 및 장치를 제공한다.

[0108] 본 개시의 다중 객체 추적 방법은 컴퓨터 장치(100)에 의해 수행되며, 글로벌 외관 모델을 이용하여, 연속 프레임들에서 감지되는 객체들의 외관 특징들을 검출하는 단계, 글로벌 상대 모션 모델을 이용하여, 감지되는 객체들의 미래 궤적들을 예측하는 단계, 및 외관 특징들과 미래 궤적들을 기반으로 감지되는 객체들을 구별하여 연계시키는 단계를 포함할 수 있다.

[0109] 다양한 실시예들에 따르면, 글로벌 외관 모델 또는 글로벌 상대 모션 모델 중 적어도 하나는, 학습 시 미리 정해진 제약 조건을 기반으로 업데이트될 수 있다.

[0110] 다양한 실시예들에 따르면, 감지되는 객체들을 연계시키는 단계는, 외관 특징들과 미래 궤적들을 기반으로 감지되는 객체들을 구별하여 연계시킬 수 있다.

[0111] 다양한 실시예들에 따르면, 글로벌 외관 모델은, 학습 시, 훈련 데이터셋에서 감지되는 객체들을 아이디(ID)들로 정의하고, 타겟인 앵커 객체, 앵커 객체와 동일한 아이디의 포지티브 객체, 및 앵커 객체와 다른 아이디의 네거티브 객체를 선택하고, 앵커 객체와 포지티브 객체의 외관 특징들에 따른 포지티브 객체 쌍들, 및 앵커 객체와 네거티브 객체의 외관 특징들에 따른 네거티브 객체 쌍들을 정의하고, 트리플렛 손실을 기반으로, 포지티브 객체 쌍들 사이의 특징 거리를 최소화하고, 네거티브 객체 쌍들 사이의 특징 거리를 최대화할 수 있다.

[0112] 다양한 실시예들에 따르면, 글로벌 상대 모션 모델은, 훈련 데이터셋에서 감지되는 객체들의 각각의 과거 궤적 및 객체들 사이의 상대 모션을 기반으로 객체들의 각각의 미래 궤적을 예측하도록, 학습될 수 있다.

[0113] 다양한 실시예들에 따르면, 글로벌 상대 모션 모델은, 과거 궤적 및 상대 모션을 기반으로 객체들의 각각의 미래 궤적을 생성하도록 구성되는 생성기, 및 예측된 미래 궤적이 훈련 데이터셋에서 확인되는 실측 궤적과 일치하는지의 여부를 판별하도록 구성되는 판별기를 포함할 수 있다.

[0114] 다양한 실시예들에 따르면, 새 객체가 나타나는 경우, 글로벌 외관 모델 및 글로벌 상대 모션 모델이 업데이트되고, 한 객체에 대해 상이한 프레임들에서 검출되는 외관 특징들의 유사도가 임계값 이하인 경우, 글로벌 외관 모델이 업데이트되고, 현재 프레임에서 예측되는 미래 궤적들의 수가 미리 정해진 수 미만이면, 글로벌 상대 모션 모델이 업데이트될 수 있다.

[0115] 본 개시의 다중 객체 추적 장치는 컴퓨터 장치(100)로서, 메모리(130), 및 메모리(130)와 연결되고, 메모리(130)에 저장된 적어도 하나의 명령을 실행하도록 구성되는 프로세서(140)를 포함하고, 프로세서(140)는, 글로벌 외관 모델을 이용하여, 연속 프레임들에서 감지되는 객체들의 외관 특징들을 검출하고, 글로벌 상대 모션 모델을 이용하여, 감지되는 객체들의 미래 궤적들을 예측하고, 외관 특징들과 미래 궤적들을 기반으로 감지되는 객체들을 구별하여 연계시키도록 구성될 수 있다.

[0116] 다양한 실시예들에 따르면, 글로벌 외관 모델 또는 글로벌 상대 모션 모델 중 적어도 하나는, 학습 시 미리 정

해진 제약 조건을 기반으로 업데이트될 수 있다.

- [0117] 다양한 실시예들에 따르면, 프로세서(140)는, 외관 특징들과 미래 궤적들을 기반으로 감지되는 객체들을 구별하여 연계시킬 수 있다.
- [0118] 다양한 실시예들에 따르면, 글로벌 외관 모델은, 학습 시, 훈련 데이터셋에서 감지되는 객체들을 아이디(ID)들로 정의하고, 타겟인 앵커 객체, 앵커 객체와 동일한 아이디의 포지티브 객체, 및 앵커 객체와 다른 아이디의 네거티브 객체를 선택하고, 앵커 객체와 포지티브 객체의 외관 특징들에 따른 포지티브 객체 쌍들, 및 앵커 객체와 네거티브 객체의 외관 특징들에 따른 네거티브 객체 쌍들을 정의하고, 트리플렛 손실을 기반으로, 포지티브 객체 쌍들 사이의 특징 거리를 최소화하고, 네거티브 객체 쌍들 사이의 특징 거리를 최대화할 수 있다.
- [0119] 다양한 실시예들에 따르면, 글로벌 상대 모션 모델은, 훈련 데이터셋에서 감지되는 객체들의 각각의 과거 궤적 및 객체들 사이의 상대 모션을 기반으로 객체들의 각각의 미래 궤적을 예측하도록, 학습될 수 있다.
- [0120] 다양한 실시예들에 따르면, 글로벌 상대 모션 모델은, 과거 궤적 및 상대 모션을 기반으로 객체들의 각각의 미래 궤적을 생성하도록 구성되는 생성기, 및 예측된 미래 궤적이 훈련 데이터셋에서 확인되는 실측 궤적과 일치하는지의 여부를 판별하도록 구성되는 판별기를 포함할 수 있다.
- [0121] 다양한 실시예들에 따르면, 프로세서(140)는, 새 객체가 나타나는 경우, 글로벌 외관 모델 및 글로벌 상대 모션 모델을 업데이트하고, 한 객체에 대해 상이한 프레임들에서 검출되는 외관 특징들의 유사도가 임계값 이하인 경우, 글로벌 외관 모델을 업데이트하고, 현재 프레임에서 예측되는 미래 궤적들의 수가 미리 정해진 수 미만이면, 글로벌 상대 모션 모델을 업데이트하도록 구성될 수 있다.
- [0122] 상술한 방법은 컴퓨터에서 실행하기 위해 컴퓨터 판독 가능한 기록 매체에 저장된 컴퓨터 프로그램으로 제공될 수 있다. 매체는 컴퓨터로 실행 가능한 프로그램을 계속 저장하거나, 실행 또는 다운로드를 위해 임시 저장하는 것일 수도 있다. 또한, 매체는 단일 또는 수개 하드웨어가 연계된 형태의 다양한 기록 수단 또는 저장수단일 수 있는데, 어떤 컴퓨터 시스템에 직접 접속되는 매체에 한정되지 않고, 네트워크 상에 분산 존재하는 것일 수도 있다. 매체의 예시로는, 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체, CD-ROM 및 DVD와 같은 광기록 매체, 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical medium), 및 ROM, RAM, 플래시 메모리 등을 포함하여 프로그램 명령어가 저장되도록 구성된 것이 있을 수 있다. 또한, 다른 매체의 예시로, 애플리케이션을 유통하는 앱 스토어나 기타 다양한 소프트웨어를 공급 내지 유통하는 사이트, 서버 등에서 관리하는 기록매체 내지 저장매체도 들 수 있다.
- [0123] 본 개시의 방법, 동작 또는 기법들은 다양한 수단에 의해 구현될 수도 있다. 예를 들어, 이러한 기법들은 하드웨어, 펌웨어, 소프트웨어, 또는 이들의 조합으로 구현될 수도 있다. 본원의 개시와 연계하여 설명된 다양한 예시적인 논리적 블록들, 모듈들, 회로들, 및 알고리즘 단계들은 전자 하드웨어, 컴퓨터 소프트웨어, 또는 양자의 조합들로 구현될 수도 있음을 통상의 기술자들은 이해할 것이다. 하드웨어 및 소프트웨어의 이러한 상호 대체를 명확하게 설명하기 위해, 다양한 예시적인 구성요소들, 블록들, 모듈들, 회로들, 및 단계들이 그들의 기능적 관점에서 일반적으로 위에서 설명되었다. 그러한 기능이 하드웨어로서 구현되는지 또는 소프트웨어로서 구현되는지의 여부는, 특정 애플리케이션 및 전체 시스템에 부과되는 설계 요구사항들에 따라 달라진다. 통상의 기술자들은 각각의 특정 애플리케이션을 위해 다양한 방식으로 설명된 기능을 구현할 수도 있으나, 그러한 구현들은 본 개시의 범위로부터 벗어나게 하는 것으로 해석되어서는 안된다.
- [0124] 하드웨어 구현에서, 기법들을 수행하는 데 이용되는 프로세싱 유닛들은, 하나 이상의 ASIC들, DSP들, 디지털 신호 프로세싱 디바이스들(digital signal processing devices; DSPD들), 프로그램가능 논리 디바이스들(programmable logic devices; PLD들), 필드 프로그램가능 게이트 어레이들(field programmable gate arrays; FPGA들), 프로세서들, 제어기들, 마이크로제어기들, 마이크로프로세서들, 전자 디바이스들, 본 개시에 설명된 기능들을 수행하도록 설계된 다른 전자 유닛들, 컴퓨터, 또는 이들의 조합 내에서 구현될 수도 있다.
- [0125] 따라서, 본 개시와 연계하여 설명된 다양한 예시적인 논리 블록들, 모듈들, 및 회로들은 범용 프로세서, DSP, ASIC, FPGA나 다른 프로그램 가능 논리 디바이스, 이산 게이트나 트랜지스터 로직, 이산 하드웨어 컴포넌트들, 또는 본원에 설명된 기능들을 수행하도록 설계된 것들의 임의의 조합으로 구현되거나 수행될 수도 있다. 범용 프로세서는 마이크로프로세서일 수도 있지만, 대안으로, 프로세서는 임의의 종래의 프로세서, 제어기, 마이크로제어기, 또는 상태 머신일 수도 있다. 프로세서는 또한, 컴퓨팅 디바이스들의 조합, 예를 들면, DSP와 마이크로프로세서, 복수의 마이크로프로세서들, DSP 코어와 연계한 하나 이상의 마이크로프로세서들, 또는 임의의 다른 구성의 조합으로서 구현될 수도 있다.

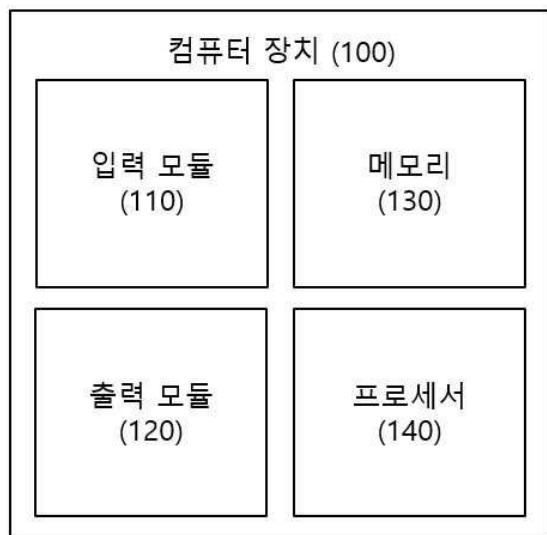
[0126] 펌웨어 및/또는 소프트웨어 구현에 있어서, 기법들은 랜덤 액세스 메모리(random access memory; RAM), 판독 전용 메모리(read-only memory; ROM), 비휘발성 RAM(non-volatile random access memory; NVRAM), PROM(programmable read-only memory), EPROM(erasable programmable read-only memory), EEPROM(electrically erasable PROM), 플래시 메모리, 콤팩트 디스크(compact disc; CD), 자기 또는 광학 데이터 스토리지 디바이스 등과 같은 컴퓨터 판독가능 매체 상에 저장된 명령들로서 구현될 수도 있다. 명령들은 하나 이상의 프로세서들에 의해 실행 가능할 수도 있고, 프로세서(들)로 하여금 본 개시에 설명된 기능의 특정 양태들을 수행하게 할 수도 있다.

[0127] 이상 설명된 실시예들이 하나 이상의 독립형 컴퓨터 시스템에서 현재 개시된 주제의 양태들을 활용하는 것으로 기술되었으나, 본 개시는 이에 한정되 지 않고, 네트워크나 분산 컴퓨팅 환경과 같은 임의의 컴퓨팅 환경과 연계하여 구현될 수도 있다. 또 나아가, 본 개시에서 주제의 양상들은 복수의 프로세싱 칩들이나 장치들에서 구현될 수도 있고, 스토리지는 복수의 장치들에 걸쳐 유사하게 영향을 받게 될 수도 있다. 이러한 장치들은 PC들, 네트워크 서버들, 및 휴대용 장치들을 포함할 수도 있다.

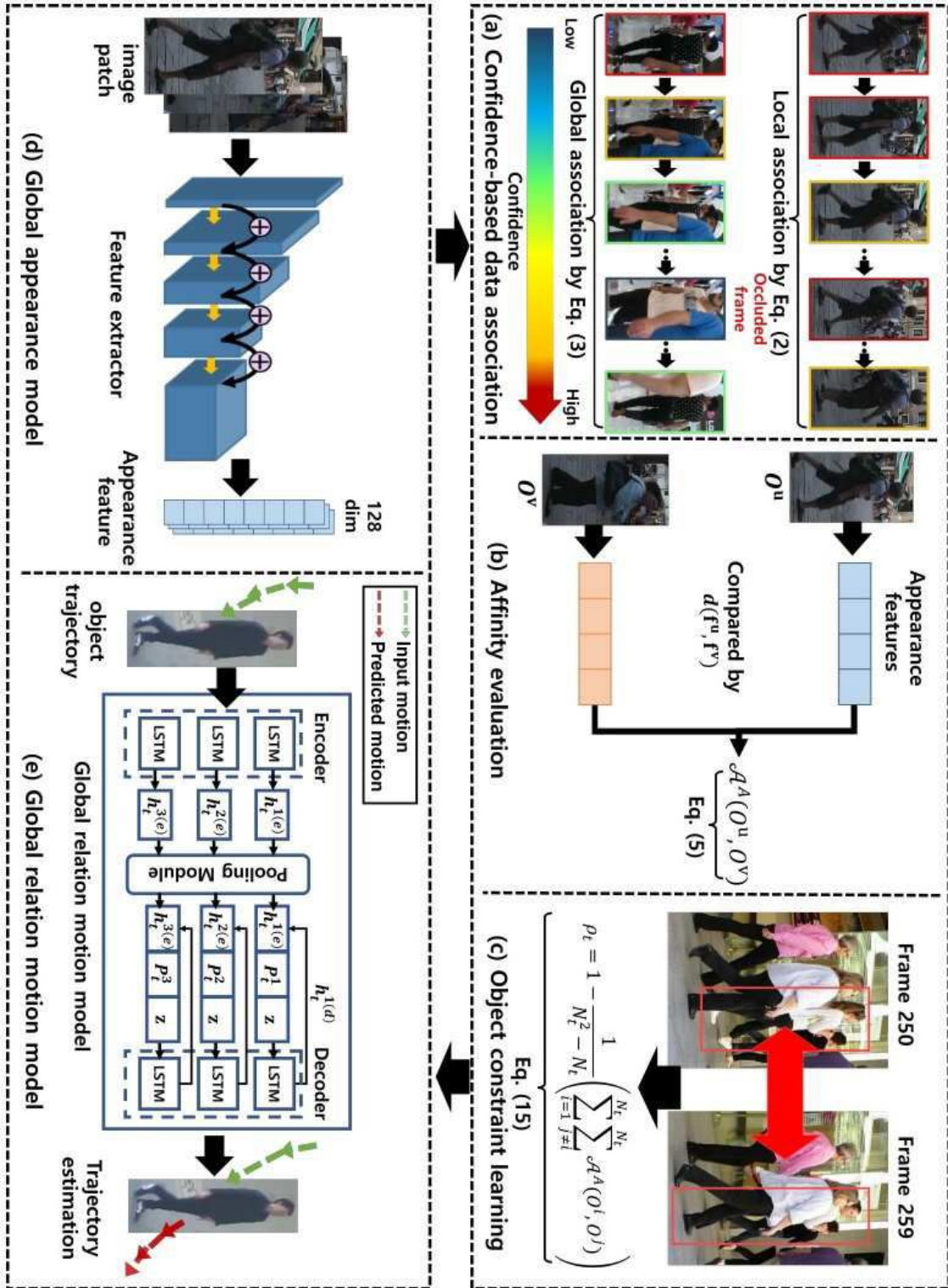
[0128] 본 개시가 일부 실시예들과 관련하여 설명되었지만, 본 개시의 발명이 속하는 기술분야의 통상의 기술자가 이해할 수 있는 본 개시의 범위를 벗어나지 않는 범위에서 다양한 변형 및 변경이 이루어질 수 있다. 또한, 그러한 변형 및 변경은 본 문서에 첨부된 청구범위 내에 속하는 것으로 생각되어야 한다.

도면

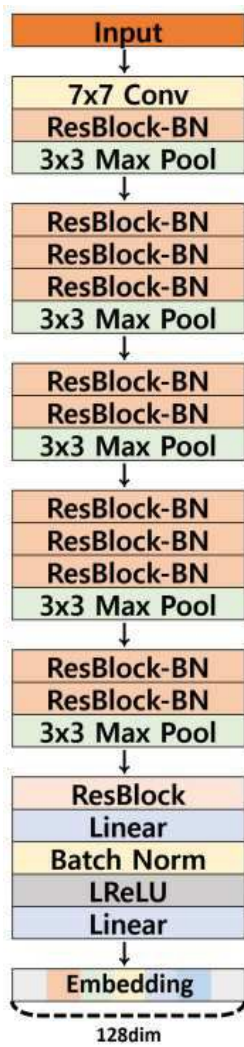
도면1



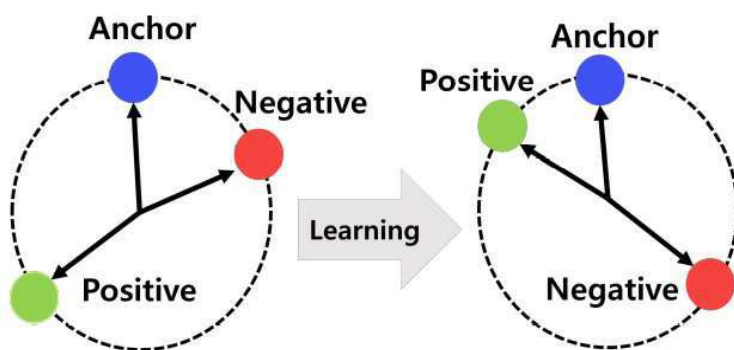
도면2



도면3



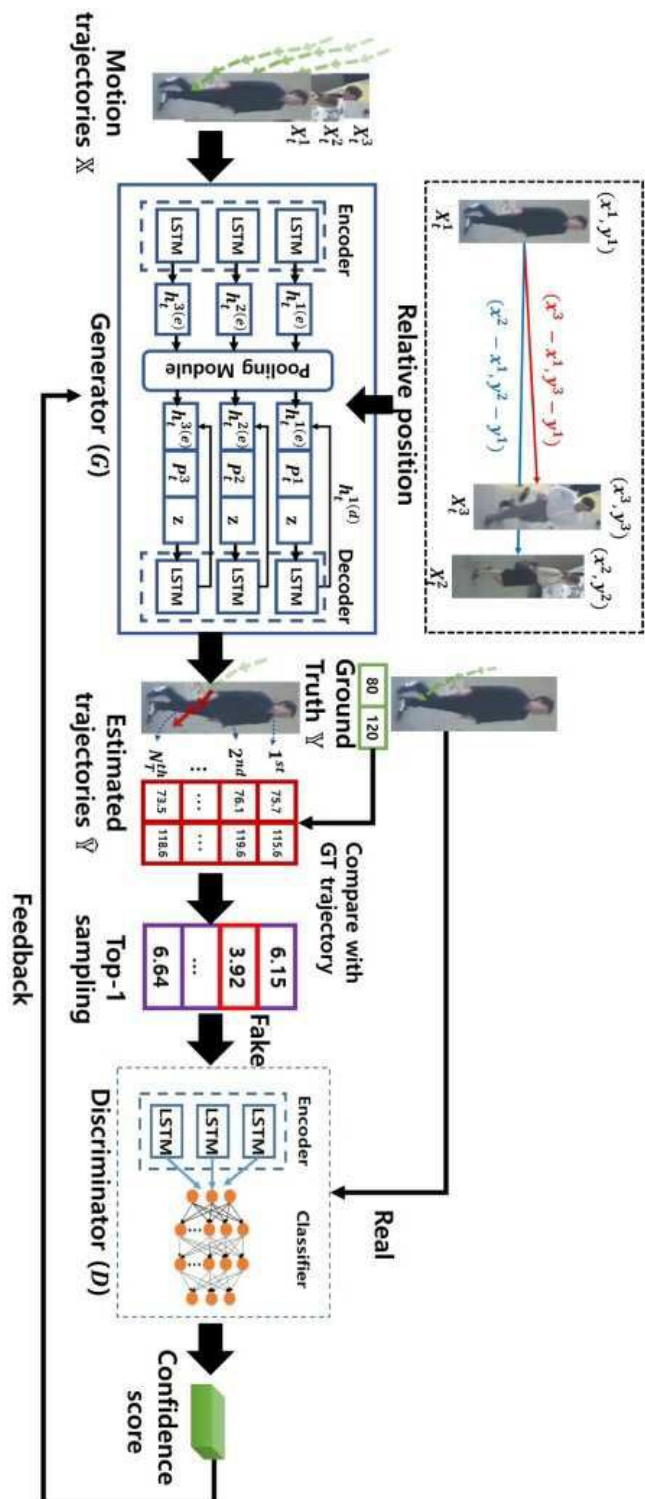
도면4

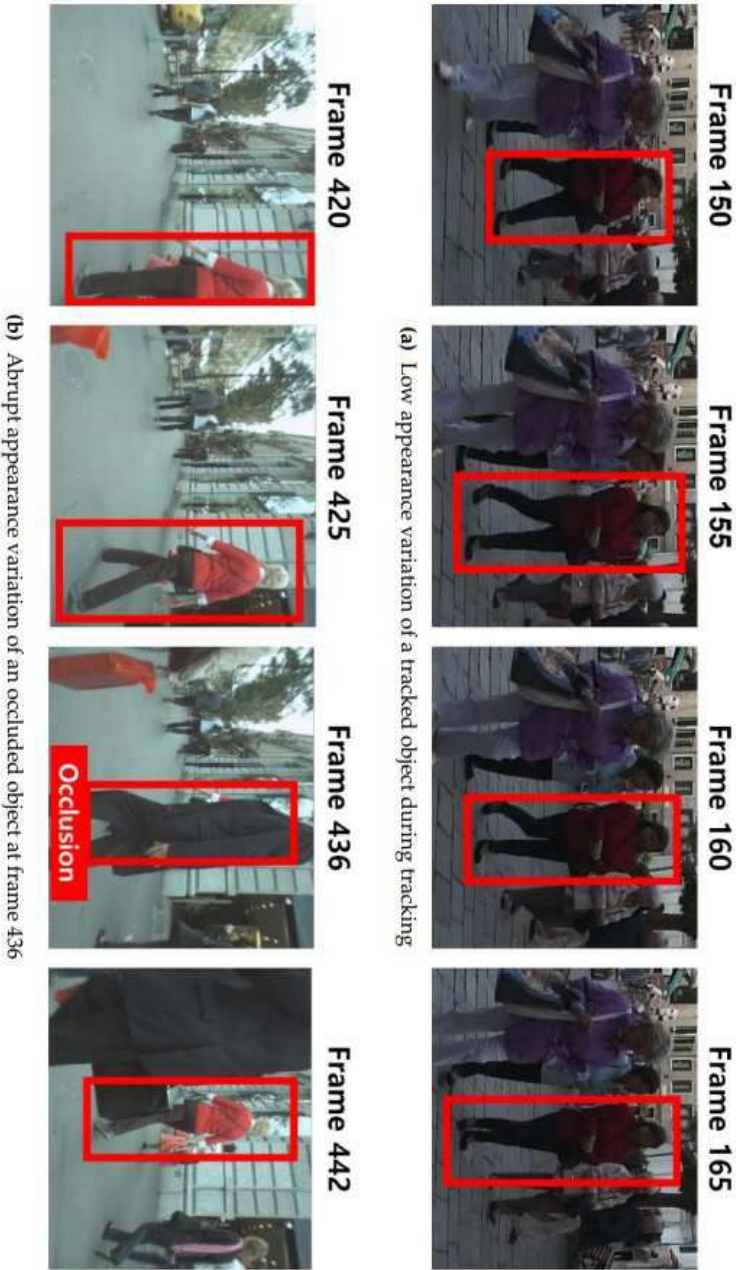


도면5



도면6





도면7

도면8

Algorithm 1: Object constraint learning for the global appearance model

Input : Tracklets $\{O^i\}_{i=1}^{N_t}$ at frame t , and N_{t-1}
Output: Updated discriminability measure ρ_t

```

1  $\rho_t \leftarrow 0, k \leftarrow 0$ 
2 Initialize an appearance dissimilarity array  $L = 0$  with  $N_t$  elements
3 for  $i = 1$  to  $N_t$  do
4   for  $j = 1$  to  $N_t$  do
5     if  $i \neq j$  then
6        $L[k] \leftarrow \mathcal{A}^A(O^i, O^j)$ 
7        $k \leftarrow k + 1$ 
8     end
9   end
10 end
11 Perform minmax normalization for  $L$ 
12 for  $i = 1$  to  $k$  do
13    $\rho_t \leftarrow \rho_t + L[i]$ 
14 end
15  $\rho_t \leftarrow 1 - \frac{\rho_t}{N_t^2 - N_t};$ 

```

도면9

Algorithm 2: The overall MOT algorithm with the proposed global models and object constraint learning

Input : A set of detections $\mathbb{Z}_t$, a set of tracklets $\{O^i\}_{i=1}^{N_t}$, and N_{t-1}
Output: Updated tracklets $\{O^i\}_{i=1}^{N_t}$

```

1 // Step1: Confidence-based data association;
2 // Local association;
3 for  $i = 1$  to  $h$  do
4     for  $j = 1$  to  $n$  do
5         Compute local association matrix  $\mathbf{S}_{h \times n}^{local} = [s_{ij}]$  by Eq. (2)
6     end
7 end
8 Optimize  $\mathbf{S}_{h \times n}^{local}$  for local association
9 // Global association;
10 for  $i = 1$  to  $l + \eta$  do
11     for  $j = 1$  to  $h + l$  do
12         Compute a global association matrix  $\mathbf{S}_{(l+\eta) \times (h+l)}^{global}$  by Eq. (3)
13     end
14 end
15 Optimize  $\mathbf{S}_{(l+\eta) \times (h+l)}^{global}$  for global association
16 // Step2: Model and confidence update;
17 Compute  $\rho_t$  using Algorithm 1
18 for  $i = 1$  to  $N_t$  do
19     if  $N_t \neq N_{t-1}$  then
20         Update  $O^i\{A\}$  by a global appearance model
21         Update  $O^i\{M\}$  by a global motion model
22     end
23     else
24         if  $\rho_t < \mu$  then
25             Update  $O^i\{A\}$  by a global appearance model
26         end
27         if  $\Delta > \Delta_{est}$  then
28             Update  $O^i\{M\}$  by a global motion model
29         end
30     end
31     Update  $O^i\{S\}$ 
32     Update the confidence of  $O^i$  by Eq. (1)
33 end

```

도면10

