(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0006971
(43) 공개일자 2021년01월19일

(51) 국제특허분류(Int. Cl.)
G06K 9/00 (2006.01) *G06K 9/46* (2006.01)
G06K 9/62 (2006.01) *G06N 20/00* (2019.01)
(52) CPC특허분류
G06K 9/00697 (2013.01)
G06F 16/783 (2019.01)
(21) 출원번호 10-2020-7035322
(22) 출원일자(국제) 2019년01월10일
심사청구일자 2020년12월08일
(85) 번역문제출일자 2020년12월08일
(86) 국제출원번호 PCT/US2019/013024
(87) 국제공개번호 WO 2020/131140
국제공개일자 2020년06월25일
(30) 우선권주장
62/781,276 2018년12월18일 미국(US)

(71) 출원인
구글 엘엘씨
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이 1600 (우:94043)
(72) 발명자
고르반 알렉산더
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이 1600 (우:94043) 구글 엘엘씨 씨/오
우 옌샹
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이 1600 (우:94043) 구글 엘엘씨 씨/오
(74) 대리인
박장원

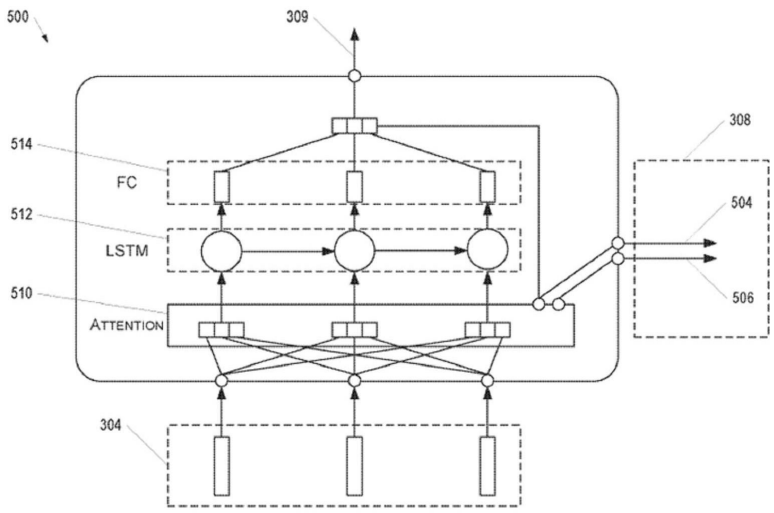
전체 청구항 수 : 총 15 항

(54) 발명의 명칭 **지오로케이션 예측을 위한 시스템 및 방법**

(57) 요 약

하나의 예시적인 실시예에서, 이미지로부터 정보를 추출하기 위한 컴퓨터 구현 방법은 이미지 시퀀스를 나타내는 데이터를 획득하는 단계를 포함하고, 이미지 시퀀스의 적어도 하나는 객체를 묘사한다. 방법은 이미지 시퀀스로부터 위치 정보를 추출하도록 트레이닝된 기계 학습 정보 추출 모델에 이미지 시퀀스를 입력하는 단계를 포함한다. 방법은 이미지 시퀀스의 입력에 응답하여 정보 추출 모델의 출력으로서 이미지 시퀀스에 묘사된 객체와 관련된 실제 위치를 나타내는 데이터를 획득하는 단계를 포함한다.

대 표 도



(52) CPC특허분류
G06K 9/4628 (2013.01)
G06K 9/6267 (2013.01)
G06N 20/00 (2019.01)
G06K 2209/27 (2013.01)

명 세 서

청구범위

청구항 1

이미지로부터 정보를 추출하기 위한 컴퓨터 구현 방법으로서, 상기 방법은,

하나 이상의 프로세서를 포함하는 컴퓨팅 시스템에서, 이미지 시퀀스를 나타내는 데이터를 획득하는 단계, 상기 이미지 시퀀스 중 적어도 하나는 객체를 묘사하고;

컴퓨팅 시스템에 의해, 이미지 시퀀스를 이미지 시퀀스로부터 위치 정보를 추출하도록 트레이닝된 기계 학습된 정보 추출 모델에 입력하는 단계; 및

컴퓨팅 시스템에 의해, 이미지 시퀀스를 입력하는 것에 응답하여 정보 추출 모델의 출력으로서, 이미지 시퀀스에 묘사된 객체와 관련된 실제 위치를 나타내는 데이터를 획득하는 단계를 포함하는 것을 특징으로 하는 이미지로부터 정보를 추출하기 위한 컴퓨터 구현 방법.

청구항 2

제1항에 있어서,

이미지 시퀀스에 묘사된 객체와 관련된 실제 위치를 나타내는 데이터를 획득하는 단계는,

컴퓨팅 시스템에 의해, 이미지 프레임 시퀀스에 묘사된 객체와 관련된 분류를 나타내는 데이터를 결정하는 단계;

컴퓨팅 시스템에 의해, 이미지 시퀀스로부터 추출된 이미지 특징 시퀀스를 나타내는 데이터에 적어도 부분적으로 기초하여, 이미지 시퀀스와 관련된 시간적 관심값(attention value) 및 공간적 관심값을 결정하는 단계; 및

컴퓨팅 시스템에 의해, 이미지 특징 시퀀스, 시간적 관심값 및 공간적 관심값에 적어도 부분적으로 기초하여, 객체와 관련된 실제 위치를 예측하는 단계를 포함하는 것을 특징으로 하는 이미지로부터 정보를 추출하기 위한 컴퓨터 구현 방법.

청구항 3

제2항 에 있어서,

이미지 시퀀스와 관련된 시간적 관심값과 공간적 관심값을 결정하는 단계는,

컴퓨팅 시스템에 의해, 약하게 지도된(weakly supervised) 객체 분류 모델에 이미지 특징 시퀀스를 입력하는 단계, 상기 객체 분류 모델은 적어도 하나의 장단기 메모리 블록을 포함하고; 및

컴퓨팅 시스템에 의해, 이미지 특징 시퀀스를 입력하는 것에 응답하여 객체 분류 모델의 출력으로서, 시간적 관심값과 공간적 관심값을 획득하는 단계를 포함하는 것을 특징으로 하는 컴퓨터 구현 방법.

청구항 4

제2항 또는 제3항에 있어서,

이미지 프레임 시퀀스에 묘사된 객체와 관련된 분류를 결정하는 단계는,

컴퓨팅 시스템에 의해, 이미지 특징 시퀀스를 약하게 지도된 객체 분류 모델에 입력하는 단계; 및

컴퓨팅 시스템에 의해, 이미지 특징 시퀀스를 입력하는 것에 응답하여 객체 분류 모델의 출력으로서, 객체와 관련된 분류를 획득하는 단계를 포함하는 것을 특징으로 하는 컴퓨터 구현 방법.

청구항 5

제4항에 있어서,

이미지 프레임 시퀀스를 나타내는 데이터는 이미지 프레임 시퀀스와 관련된 적어도 하나의 분류 라벨(label)을

포함하고, 그리고 상기 방법은,

컴퓨팅 시스템에 의해, 이미지 시퀀스와 관련된 적어도 하나의 분류 라벨에 적어도 부분적으로 기초하여, 객체 분류 모델에 의해 출력된 분류와 관련된 손실을 결정하는 단계; 및

컴퓨팅 시스템에 의해, 결정된 손실에 적어도 부분적으로 기초하여 객체 분류 모델을 트레이닝하는 단계를 더 포함하는 것을 특징으로 하는 컴퓨터 구현 방법.

청구항 6

제2항, 제3항, 제4항 또는 제5항에 있어서,

객체와 관련된 실제 위치를 예측하는 단계는,

컴퓨팅 시스템에 의해, 이미지 특징 시퀀스, 시간적 관심값 및 공간적 관심값을 프레임 레벨 위치-특징 추출 모델에 입력하는 단계;

컴퓨팅 시스템에 의해, 이미지 특징 시퀀스, 시간적 관심값 및 공간적 관심값을 입력하는 것에 응답하여 프레임 레벨 위치-특징 추출 모델의 출력으로서, 객체와 관련된 하나 이상의 위치 특징을 포함하는 위치 특징 시퀀스를 나타내는 데이터를 획득하는 단계;

컴퓨팅 시스템에 의해, 위치 특징 시퀀스를 프레임 레벨 위치 예측 모델에 입력하는 단계;

컴퓨팅 시스템에 의해, 위치 특징 시퀀스를 입력하는 것에 응답하여 프레임 레벨 위치 예측 모델의 출력으로서, 객체와 관련된 카메라 좌표 공간의 좌표를 나타내는 데이터를 획득하는 단계; 및

컴퓨팅 시스템에 의해, 카메라 좌표 공간의 좌표 및 객체와 관련된 카메라 포즈 데이터에 적어도 부분적으로 기초하여 객체와 관련된 실제 좌표를 결정하는 단계를 포함하는 것을 특징으로 하는 컴퓨터 구현 방법.

청구항 7

제6항에 있어서,

컴퓨팅 시스템에 의해, 객체를 묘사하는 이미지 시퀀스의 다수의 이미지에 걸쳐 객체와 관련된 좌표 사이의 변화에 적어도 부분적으로 기초하여 위치 일관성 손실을 결정하는 단계; 및

컴퓨팅 시스템에 의해, 위치 일관성 손실에 적어도 부분적으로 기초하여 프레임 레벨 위치 예측 모델을 트레이닝하는 단계를 더 포함하는 것을 특징으로 하는 컴퓨터 구현 방법.

청구항 8

제6항 또는 제7항에 있어서, .

컴퓨팅 시스템에 의해, 객체를 묘사하는 이미지 시퀀스의 다수의 이미지에 걸쳐 결정된 외관 특징들 사이의 변화에 적어도 부분적으로 기초하여 외관 일관성 손실을 결정하는 단계; 및

컴퓨팅 시스템에 의해, 외관 일관성 손실에 적어도 부분적으로 기초하여 프레임-레벨 위치 예측 모델을 트레이닝하는 단계를 더 포함하는 것을 특징으로 하는 컴퓨터 구현 방법.

청구항 9

제6항, 제7항 또는 제8항에 있어서,

컴퓨팅 시스템에 의해, 객체와 관련된 카메라 좌표 공간의 좌표 및 객체를 묘사하는 이미지 시퀀스의 다수의 이미지에 걸쳐 객체와 관련된 공간적 관심에 적어도 부분적으로 기초하여 에이밍(aiming) 손실을 결정하는 단계; 및

컴퓨팅 시스템에 의해, 에이밍 손실에 적어도 부분적으로 기초하여 프레임 레벨 위치 예측 모델을 트레이닝하는 단계를 더 포함하는 것을 특징으로 하는 컴퓨터 구현 방법.

청구항 10

제6항, 제7항, 제8항 또는 제9항에 있어서,

컴퓨팅 시스템에 의해, 객체와 관련된 실제 좌표 및 객체를 묘사하는 이미지 시퀀스를 캡처하는데 사용된 카메라와 관련된 시야(field-of-view)에 적어도 부분적으로 기초하여 시야 손실을 결정하는 단계; 및

컴퓨팅 시스템에 의해, 시야 손실에 적어도 부분적으로 기초하여 프레임 레벨 위치 예측 모델을 트레이닝하는 단계를 더 포함하는 것을 특징으로 하는 컴퓨터 구현 방법.

청구항 11

선행하는 청구항 중 어느 한 항에 있어서,

이미지 시퀀스는 그 이미지 시퀀스의 다수의 이미지에 걸쳐 복수의 객체를 묘사하고, 정보 추출 모델의 출력은 이미지 시퀀스에 묘사된 복수의 객체와 관련된 실제 위치를 나타내는 데이터를 포함하는 것을 특징으로 하는 컴퓨터 구현 방법.

청구항 12

이미지 시퀀스에 묘사된 객체와 관련된 실제 위치를 나타내는 데이터를 결정하기 위해 정보 추출 모델을 트레이닝하기 위한 컴퓨터 구현 방법으로서, 상기 정보 추출 모델은,

이미지-특징 추출 모델;

약하게 지도된 객체 분류 모델;

지오로케이션 예측 모델을 포함하고,

상기 방법은, 하나 이상의 프로세서를 포함하는 컴퓨팅 시스템에서:

객체를 묘사하는 이미지 시퀀스 중 적어도 하나에서, 노이즈 분류를 이용하여 이미지 시퀀스를 나타내는 데이터를 획득하는 단계;

이미지 시퀀스를 입력하는 것에 응답하여 이미지-특징 추출 모델에 의해, 이미지 특징 시퀀스를 출력하는 단계;

이미지 특징 시퀀스를 입력하는 것에 응답하여 객체 분류 모델에 의해, 이미지 시퀀스와 관련된 하나 이상의 분류 라벨을 포함하는 분류 데이터를 출력하는 단계, 상기 분류 데이터는 이미지 특징 시퀀스와 관련된 하나 이상의 시간적 관심값 및 하나 이상의 공간적 관심값에 적어도 부분적으로 기초하여 결정되고, 상기 하나 이상의 시간적 관심값과 하나 이상의 공간적 관심값는 객체 분류 모델에 의해 결정되며;

이미지 시퀀스와 관련된 분류 데이터 및 노이즈 분류에 적어도 부분적으로 기초하여 객체 분류 모델을 트레이닝하는 단계;

이미지 특징 시퀀스, 하나 이상의 시간적 관심값 및 하나 이상의 공간적 관심값을 입력하는 것에 응답하여 지오로케이션 예측 모델에 의해, 이미지 시퀀스에 묘사된 객체와 관련된 실제 세계 위치를 출력하는 단계; 및

적어도 이미지 특징 시퀀스, 시간적 관심값 및 공간적 관심값을 사용하여 지오로케이션 예측 모델을 트레이닝하는 단계를 포함하는 것을 특징으로 하는 컴퓨터 구현 방법.

청구항 13

이미지로부터 정보를 추출하기 위한 컴퓨터 구현 방법으로서, 상기 방법은,

하나 이상의 프로세서를 포함하는 컴퓨팅 시스템에서, 하나 이상의 이미지를 나타내는 데이터를 획득하는 단계, 상기 하나 이상의 이미지 중 적어도 하나는 객체를 묘사하고;

컴퓨팅 시스템에 의해, 하나 이상의 이미지로부터 위치 정보를 추출하도록 트레이닝된 기계 학습된 정보 추출 모델에 하나 이상의 이미지를 입력하는 단계; 및

컴퓨팅 시스템에 의해, 하나 이상의 이미지를 입력하는 것에 응답하여 정보 추출 모델의 출력으로서, 하나 이상의 이미지에 묘사된 객체와 관련된 실제 위치를 나타내는 데이터를 획득하는 단계를 포함하는 것을 특징으로 하는 컴퓨터 구현 방법.

청구항 14

컴퓨팅 시스템으로서, 상기 시스템은,

하나 이상의 프로세서;

하나 이상의 기계 학습 정보 추출 모델; 및

하나 이상의 프로세서에 의해 실행될 때 시스템으로 하여금 제1항 내지 제13항의 방법을 수행하게 하는 명령들이 저장된 컴퓨터 판독 가능 매체를 포함하는 것을 특징으로 하는 컴퓨팅 시스템.

청구항 15

하나 이상의 프로세서에 의해 실행될 때 하나 이상의 프로세서로 하여금 제1항 내지 제13항 중 어느 한 항에 따른 동작들을 수행하게 하는 하나 이상의 기계 학습 정보 추출 모델 및 컴퓨터 판독 가능 명령들을 저장하는 하나 이상의 유형의 비-일시적 컴퓨터 판독 가능 매체.

발명의 설명

기 술 분 야

[0001] 본 발명은 일반적으로 하나 이상의 객체에 대한 실제 위치(예를 들어, 지오로케이션)를 예측하는 것에 관한 것이다. 보다 구체적으로, 본 발명은 객체(들)에 대한 실제 위치(들)를 예측하기 위해 비지도 방식으로 트레이닝될 수 있는 정보 추출 모델에 관한 것이다.

배 경 기 술

[0002] 기계 학습된 모델을 사용하여 이미지로부터 데이터를 추출하기 위한 주요 장애물 중 하나는 이러한 모델을 트레이닝시키기에 충분한 데이터를 얻는데 많은 비용이 든다는 점이다. 임의의 객체에 대한 실제 좌표를 포함하는 지상 실측 데이터를 얻는 것은 많은 시간이 소요되는 프로세스이다. 대부분의 경우, 실제(real-world) 위치를 예측하기 위해 신경망을 트레이닝하는데 충분한 지상 실측 데이터를 얻는 것은 비현실적이다. 다른 문제는 이미지에서 추출되어야 하는 다양한 유형의 데이터가 빠르게 증가한다는 점이다.

발명의 내용

[0003] 본 발명의 실시예의 양태 및 장점은 다음의 설명에서 부분적으로 설명되거나, 설명으로부터 학습될 수 있거나 실시예의 실행을 통해 학습될 수 있다.

[0004] 본 발명의 하나의 예시적인 양태는 이미지에서 정보를 추출하기 위한 컴퓨터 구현 방법에 관한 것이다. 방법은 하나 이상의 프로세서를 포함하는 컴퓨팅 시스템에서, 이미지 시퀀스를 나타내는 데이터를 획득하는 단계를 포함하며, 이미지 시퀀스 중 적어도 하나는 객체를 묘사한다. 방법은 컴퓨팅 시스템에 의해, 이미지 시퀀스로부터 위치 정보를 추출하도록 트레이닝된 기계 학습 정보 추출 모델에 이미지 시퀀스를 입력하는 단계를 포함한다. 방법은 컴퓨팅 시스템에 의해, 이미지 시퀀스를 입력하는 것에 응답하여 정보 추출 모델의 출력으로서 이미지 시퀀스에 묘사된 객체와 관련된 실제 위치를 나타내는 데이터를 획득하는 단계를 포함한다.

[0005] 본 발명의 다른 예시적인 양태는 컴퓨팅 시스템에 관한 것이다. 컴퓨팅 시스템은 하나 이상의 프로세서, 하나 이상의 기계 학습 정보 추출 모델 및 하나 이상의 프로세서에 의해 실행될 때 컴퓨팅 시스템으로 하여금 동작들을 수행하게 하는 명령들을 집합적으로 저장하는 하나 이상의 유형의 비-일시적 컴퓨터 판독 가능 매체를 포함한다. 동작들은 이미지 시퀀스를 나타내는 데이터를 획득하는 단계를 포함하며, 이미지 시퀀스 중 적어도 하나는 객체를 묘사한다. 동작들은 이미지 시퀀스로부터 위치 정보를 추출하도록 트레이닝된 기계 학습 정보 추출 모델에 이미지 시퀀스를 입력하는 단계를 포함한다. 동작은 이미지 시퀀스를 입력하는 것에 응답하여 정보 추출 모델의 출력으로서, 이미지 시퀀스에 묘사된 객체와 관련된 실제 위치를 나타내는 데이터를 획득하는 단계를 포함한다.

[0006] 본 발명의 다른 예시적인 양태는 하나 이상의 프로세서에 의해 실행될 때 하나 이상의 프로세서로 하여금 동작들을 수행하게 하는 하나 이상의 기계 학습 정보 추출 모델 및 컴퓨터 판독 가능 명령들을 저장하는 하나 이상의 유형의 비-일시적 컴퓨터 판독 가능 매체에 대한 것이다. 동작들은 이미지 시퀀스를 나타내는 데이터를 획득하는 단계를 포함하며, 이미지 시퀀스 중 적어도 하나는 객체를 묘시한다다. 동작들은 이미지 시퀀스로부터 위치 정보를 추출하도록 트레이닝된 기계 학습 정보 추출 모델에 이미지 시퀀스를 입력하는 단계를 포함한다. 동작들은 이미지 시퀀스를 입력하는 것에 응답하여 정보 추출 모델의 출력으로서, 이미지 시퀀스에 묘사된 객체와 관련된 실제 위치를 나타내는 데이터를 획득하는 단계를 포함한다.

[0007] 본 발명의 다른 양태는 다양한 시스템, 장치, 비-일시적 컴퓨터 판독 가능 매체, 사용자 인터페이스 및 전자 디바이스에 관한 것이다.

[0008] 본 발명의 다양한 실시예의 이들 및 다른 특징, 양태 및 이점은 다음의 설명 및 첨부된 청구 범위를 참조하여 더 잘 이해될 것이다. 본 명세서에 포함되어 그 일부를 구성하는 첨부된 도면은 본 발명의 예시적인 실시예를 나타내고, 상세한 설명과 함께 관련 원리를 설명하는 역할을 한다.

도면의 간단한 설명

[0009] 본 기술의 당업자에게 지시된 실시예들의 상세한 논의는 첨부된 도면들을 참조하는 명세서에 기재되어 있다.

도 1a-1c는 본 발명의 예시적인 실시예에 따른 이미지 시퀀스로부터 위치 정보를 추출할 수 있는 예시적인 컴퓨팅 시스템/디바이스의 블록도를 도시한다.

도 2는 본 발명의 예시적인 실시예에 따른 예시적인 정보 추출 모델의 블록도를 도시한다.

도 3은 본 발명의 예시적인 실시예에 따른 예시적인 정보 추출 모델의 블록도를 도시한다.

도 4는 본 발명의 내용의 예시적인 실시예에 따른 예시적인 지오로케이션 예측 모델의 블록도를 도시한다.

도 5는 본 발명의 예시적인 실시예에 따른 예시적인 객체 분류 모델의 블록도를 도시한다.

도 6은 본 발명의 예시적인 실시예에 따른 예시적인 위치-피처 추출 모델의 블록도를 도시한다.

도 7은 본 발명의 예시적인 실시예에 따른 예시적인 위치 예측 모델의 블록도를 도시한다.

도 8은 본 발명의 예시적인 실시예에 따른 이미지 시퀀스로부터 위치 정보를 추출하기 위한 예시적인 방법의 흐름도를 도시한다.

도 9는 본 발명의 예시적인 실시예에 따른 객체 분류 모델을 트레이닝하기 위한 예시적인 방법의 흐름도를 도시한다.

도 10은 본 발명의 예시적인 실시예에 따른 위치 예측 모델을 트레이닝하기 위한 예시적인 방법의 흐름도를 도시한다.

복수의 도면에 걸쳐 반복되는 참조 번호는 다양한 구현에서 동일한 피처를 식별하기 위한 것이다.

발명을 실시하기 위한 구체적인 내용

[0010] 개요

[0011] 본 발명과 일치하는 시스템 및 방법은 복수의 이미지에 묘사된 하나 이상의 객체(예를 들어, 거리 표지판)에 대한 예측된 실제 위치(예를 들어, 위도 및 경도)를 결정하는데 사용될 수 있는 정보 추출 모델을 포함할 수 있다. 정보 추출 모델은 노이즈 분류를 갖는 이미지 데이터를 사용하여 하나 이상의 객체의 실제 위치를 예측하도록 트레이닝될 수 있다. 다량의 노이즈 라벨링된(noisy labeled) 데이터와 극히 많은 양의 라벨링되지 않은 데이터를 활용하는 기능을 가능하게 함으로써, 본 발명은 보다 많은 애플리케이션을 위한 모델을 더 빠르고 저렴하게 개발할 수 있다. 본 발명은 또한 충분한 지상 실측 위치 데이터가 이용될 수 없는 많은 새로운 애플리케이션을 가능하게 할 수 있다. 부가적으로, 정보 추출 모델은 (예를 들어, 회귀의 목표값으로) 지상 실측(ground truth) 위치 데이터 없이도 학습될 수 있으므로, 정보 추출 모델은 객체를 분류하는데 약한 지도(weak supervision)만으로 객체의 실제 위치를 예측하기 위해 비지도(unsupervised) 방식으로 종단간 모델로 트레이닝될 수 있다. 수 많은 실험을 통해 본 발명의 정보 추출 모델이 작동하고, 동일한 트레이닝 데이터와 테스트 데이터 세트를 사용하지만 모델 트레이닝시 지상 실측 라벨, 객체 경계 박스 및 강력한 지도없이도 기존의 완전한 지도 모델과 비교할 수 있는 수준의 정확도에 도달할 수 있음이 입증되었다.

[0012] 본 발명의 양태에 따르면, 하나 이상의 프로세서를 포함하는 컴퓨팅 시스템은 정보 추출 모델을 포함하는 개시된 기술의 양태를 구현하는 것을 돕는데 사용될 수 있다. 일부 구현에서, 컴퓨팅 시스템은 이미지 데이터를 획득할 수 있다. 이미지 데이터는 예를 들어 이미지 프레임 시퀀스와 같은 복수의 이미지를 포함할 수 있다. 이미지 프레임 시퀀스는 장면에서 하나 이상의 객체(예를 들어, 도로 표지판)를 묘사할 수 있다. 예를 들어, 이미지 프레임 시퀀스에 있는 하나 이상의 이미지 프레임은 하나 이상의 객체를 묘사할 수 있다. 다른 예로서, 이미지 프레임 시퀀스에 있는 복수의 이미지 프레임은 동일한 객체를 묘사할 수 있다. 동일한 객체를 묘사하는 복수의

이미지 프레임은 이미지 프레임 시퀀스에 있는 연속적인 이미지 프레임 또는 비-연속적인 이미지 프레임일 수 있다. 일부 구현에서, 이미지 프레임 시퀀스는 거리를 횡단하는 차량의 관점에서 거리에 근접한 장면을 묘사할 수 있다. 일부 구현에서, 이미지 프레임 시퀀스의 하나 이상의 이미지 프레임은 비디오의 하나 이상의 프레임 또는 다른 유형의 모션 캡처에 대응할 수 있다.

[0013] 일부 구현에서, 이미지 데이터는 이미지 프레임 시퀀스와 관련된 분류를 포함할 수 있다. 예를 들어, 이미지 데이터에는 이미지 프레임 시퀀스와 관련된 단일 분류 라벨(label)이 포함할 수 있다. 대안적으로, 이미지 데이터에는 이미지 프레임 시퀀스의 하나 이상의 이미지 프레임과 관련된 하나 이상의 분류 라벨이 포함될 수 있다. 아래에서 더 설명되는 바와 같이, 본 발명의 양태에 따르면, 시스템은 정보 추출 모델을 사용하여 노이즈 분류(예를 들어, 이미지 프레임 시퀀스와 관련된 단일 분류 라벨)를 사용하는 이미지 프레임 시퀀스에 적어도 부분적으로 기초하여 객체의 실제 위치를 나타내는 데이터를 획득할 수 있다.

[0014] 일부 구현에서, 이미지 데이터는 카메라 포즈 데이터를 포함할 수 있다. 카메라 포즈 데이터는 이미지 프레임 시퀀스에서 하나 이상의 이미지 프레임을 캡처하는데 사용되는 카메라의 실제 위치 및/또는 방향(orientation)을 나타낼 수 있다. 예를 들어, 카메라 포즈 데이터에는 4×4 카메라 대 실제 투영 행렬이 포함될 수 있다.

[0015] 본 발명의 양태에 따르면, 시스템은 지오로케이션(geolocation, 지리 위치) 데이터를 생성할 수 있다. 지오로케이션 데이터는 이미지 프레임 시퀀스에 묘사된 하나 이상의 객체(예를 들어, 도로 표지판)에 대한 예측된 실제 위치(예를 들어, 위도 및 경도)를 포함할 수 있다. 정보 추출 모델은 이미지 데이터를 수신하고, 이미지 데이터를 수신하는 것에 응답하여 지오로케이션 데이터를 출력하도록 구성될 수 있다.

[0016] 시스템은 정보 추출 모델에 이미지 데이터를 입력하고, 이미지 데이터를 입력하는 것에 응답하여 정보 추출 모델의 출력으로서 지오로케이션 데이터를 획득할 수 있다. 시스템은 지오로케이션 데이터를 사용하여 예를 들어 도로 표지판에 해당하는 도로 구간(예를 들어, 도로 표지판의 위도 및 경도 또는 그 부근에 있는 도로 구간)을 식별할 수 있다. 예를 들어, 이미지 데이터는 속도 제한 표지판을 묘사하는 이미지 프레임 시퀀스를 포함할 수 있다. 시스템은 정보 추출 모델에 이미지 데이터를 입력하고, 정보 추출 모델의 출력으로 속도 제한 표지판의 예측된 실제 위치를 포함하는 지오로케이션 데이터를 획득할 수 있다. 시스템은 지오로케이션 데이터를 사용하여 속도 제한 표지판에 해당하는 도로 구간(예를 들어, 속도 제한 표지판의 예측된 실제 좌표에 있거나 그 근처에 있는 도로 구간)을 식별할 수 있다.

[0017] 본 발명의 양태에 따르면, 정보 추출 모델은 예를 들어 이미지-특징 추출 모델, 객체 분류 모델 및 지오로케이션 예측 모델과 같은 복수의 기계 학습 모델을 포함할 수 있다. 일부 구현에서, 정보 추출 모델 및/또는 그 정보 추출 모델에 포함된 복수의 기계 학습 모델(예를 들어, 이미지-특징 추출 모델, 객체 분류 모델, 지오로케이션 예측 모델 등)은 비선형 모델 및/또는 선형 모델을 포함하는 신경망(예를 들어, 심층 신경망) 또는 다른 유형의 기계 학습 모델과 같은 다양한 기계 학습 모델일 수도 있고 이를 포함할 수 있다. 신경망에는 피드-포워드 신경망, 순환 신경망(예를 들어, 장단기 메모리 순환 신경망), 컨볼루션 신경망 또는 다른 형태의 신경망이 포함될 수 있다.

[0018] 일부 구현에서, 이미지-특징 추출 모델은 이미지 프레임 시퀀스내의 하나 이상의 이미지 프레임으로부터 추출된 하나 이상의 이미지 특징(feature)을 포함하는 이미지-특징 데이터를 생성할 수 있다. 이미지-특징 추출 모델은 이미지 프레임 시퀀스(예를 들어, 이미지 데이터)를 나타내는 데이터를 수신하여 이미지 프레임 시퀀스를 수신하는 것에 응답하여 이미지-특징 데이터를 출력하도록 구성될 수 있다. 아래에서 더 설명되는 바와 같이, 본 발명의 양태에 따라, 이미지-특징 데이터 내의 하나 이상의 이미지 특징은 이미지 프레임 시퀀스에 묘사된 하나 이상의 객체를 식별 및/또는 분류하는데 사용될 수 있다. 일부 구현에서, 이미지 특징 데이터는 이미지 특징 시퀀스를 포함할 수 있다. 예를 들어, 이미지 특징 데이터 내의 하나 이상의 이미지 특징은 이미지-특징 임베딩 시퀀스로 구성될 수 있다. 이미지-특징 임베딩 시퀀스 내의 각 이미지-특징 임베딩은 이미지 프레임 시퀀스내의 이미지 프레임에 대응할 수 있으며, 각 이미지-특징 임베딩은 대응하는 이미지 프레임에서 추출된 하나 이상의 이미지 특성을 나타낼 수 있다. 일부 구현에서, 이미지-특징 추출 모델은 예를 들어 Inception v2, 임의의 SoTA 이미지 분류 네트워크(또는 하단 부분) 등과 같은 컨벌루션 신경망(CNN)을 포함할 수 있다.

[0019] 시스템은 이미지 프레임 시퀀스(예를 들어, 이미지 데이터)를 나타내는 데이터를 이미지-특징 추출 모델에 입력하고, 이미지 프레임 시퀀스를 입력하는 것에 응답하여 이미지-특징 추출 모델의 출력으로서 이미지 특징 시퀀스(예를 들어, 이미지-특징 데이터)를 획득할 수 있다. 아래에서 더 설명되는 바와 같이, 시스템은 이미지-특징 데이터에 적어도 부분적으로 기초하여 지오로케이션 데이터를 결정하기 위해 정보 추출 모델을 사용할 수 있다.

[0020] 일부 구현에서, 객체 분류 모델은 분류 데이터 및 관심값(attention value) 데이터를 생성할 수 있다. 객체 분류 모델은 이미지 특징 시퀀스를 나타내는 데이터(예를 들어, 이미지-특징 추출 모델에 의해 출력된 이미지-특징 데이터)를 수신하고, 이미지 특징 시퀀스 및 관련 이미지-특징 임베딩을 수신하는 것에 응답하여 분류 데이터 및 관심값 데이터를 출력하도록 구성될 수 있다. 일부 구현에서, 객체 분류 모델은 약하게 지도된 순환 신경망(RNN)을 포함할 수 있다.

[0021] 분류 데이터는 이미지 프레임 시퀀스에 묘사된 하나 이상의 객체와 관련된 분류를 나타낼 수 있다. 예를 들어, 객체 분류 모델은 이미지 특징 시퀀스에 적어도 부분적으로 기초하여 하나 이상의 객체를 식별하고, 하나 이상의 식별된 객체와 관련된 하나 이상의 분류 라벨을 결정할 수 있다. 하나 이상의 객체는 이미지 프레임 시퀀스 내에 잇는 이미지 프레임의 일부 또는 전체에 묘사될 수 있다. 다른 예로서, 객체 분류 모델은 속도 제한 표지판을 묘사하는 이미지 프레임 시퀀스를 포함하는 이미지 데이터로부터 추출된 이미지 특징 시퀀스를 나타내는 데이터를 수신할 수 있다. 이미지 특징 시퀀스를 나타내는 데이터를 수신하는 것에 응답하여, 객체 분류 모델은 속도 제한 표지에 대응하는 속도 제한값을 나타내는 분류 라벨을 포함하는 분류 데이터를 출력할 수 있다.

[0022] 관심값 데이터는 예를 들어, 분류된 객체가 특정 프레임의 특정 픽셀에 존재할 확률을 나타낼 수 있다. 관심값 데이터는 이미지 특징 시퀀스와 관련된 하나 이상의 시간적(temporal) 관심값 및 하나 이상의 공간적(spatial) 관심값을 포함할 수 있다. 예를 들어, 이미지-특징 데이터는 이미지 특징 시퀀스를 나타내는 이미지-특징 임베딩 시퀀스를 포함할 수 있다. 객체 분류 모델은 이미지-특징 임베딩 시퀀스에서 각 이미지-특징 임베딩에 대한 시간적 관심값 및 공간적 관심값을 결정할 수 있다. 각 이미지-특징 임베딩에 대한 시간적 관심값 및 공간적 관심값은 분류된 객체가 이미지-특징 임베딩에 대응하는 이미지 프레임의 특정 픽셀에 존재할 확률을 나타낼 수 있다. 추가적으로 또는 대안적으로, 객체 분류 모델은 이미지 특징 시퀀스(예를 들어, 이미지-특징 임베딩 시퀀스)에 대한 단일 시간적 관심값 및 단일 공간적 관심값을 결정할 수 있다.

[0023] 시스템은 이미지 특징 시퀀스(예를 들어, 이미지-특징 데이터)를 나타내는 데이터를 객체 분류 모델에 입력하고, 이미지 특징 시퀀스를 입력하는 것에 응답하여 객체 분류 모델의 출력으로서 분류 데이터 및 관심값 데이터를 획득할 수 있다. 아래에서 더 설명하는 바와 같이, 시스템은 분류 데이터 및 관심값 데이터에 적어도 부분적으로 기초하여 지오로케이션 데이터를 결정하기 위해 정보 추출 모델을 사용할 수 있다.

[0024] 일부 구현에서, 객체 분류 모델은 시공간적 관심 메커니즘을 갖는 장단기 메모리(LSTM)를 포함할 수 있다. 시공간적 관심 메커니즘은 관심값 데이터를 결정하고 약한 지도만을 사용하여 객체 분류 모델을 효과적으로 트레이닝하는데 사용될 수 있다. 예를 들어, 객체 분류 모델은 그 객체 분류 모델에 입력된 이미지 특징 시퀀스(예를 들어, 이미지 특징 데이터)에 기초하여 프레임 당 임베딩을 각각 출력하는 복수의 LSTM 블록을 포함할 수 있다. 객체 분류 모델은 시간적 관심을 사용하여, 객체 분류 모델의 출력을 결정하기 위해 LSTM 블록에 의해 생성된 프레임 당 임베딩에 가중치를 부여할 수 있다. 이러한 방식으로, 출력의 기울기는 동일한 시간 단계에서 시간적 관심의 대응하는 가중치에 비례하여 각 LSTM 블록 사이에 동시에 배포된다.

[0025] 일부 구현에서, 객체 분류 모델은 객체 분류 모델에 의해 출력된 분류 데이터와 관련된 손실에 적어도 부분적으로 기초하여 트레이닝될 수 있다. 예를 들어, 시스템은 분류 데이터내의 하나 이상의 분류 라벨 및 이미지 데이터내의 이미지 프레임 시퀀스와 관련된 분류(예를 들어, 이미지 프레임 시퀀스와 관련된 단일 분류 라벨)에 적어도 부분적으로 기초하여 소프트맥스 교차 엔트로피 손실(softmax cross entropy loss)을 결정할 수 있다. 시스템은 결정된 소프트 맥스 교차 엔트로피를 사용하여 객체 분류 모델을 트레이닝할 수 있다.

[0026] 일부 구현에서, 정보 추출 모델은 지오로케이션 데이터를 생성할 수 있는 지오로케이션 예측 모델을 포함할 수 있다. 지오로케이션 예측 모델은 이미지 특징 시퀀스를 나타내는 데이터(예를 들어, 이미지-특징 추출 모델에 의해 출력된 이미지-특징 데이터), 이미지 특징 시퀀스에 대응하는 이미지 프레임 시퀀스를 캡처하는데 사용되는 하나 이상의 카메라와 관련된 위치 위치 및/또는 방향을 나타내는 데이터(예를 들어, 이미지 데이터내의 카메라 포즈 데이터), 및 이미지 특징 시퀀스와 관련된 관심값를 나타내는 데이터(예를 들어, 객체 분류 모델에 의해 생성된 관심값 데이터)를 수신하도록 구성될 수 있다. 지오로케이션 예측 모델은 이미지 특징 시퀀스, 카메라 위치 및/또는 방향 정보, 그리고 이미지 특징 시퀀스와 관련된 관심값들을 수신하는 것에 응답하여 지오로케이션 데이터를 출력하도록 구성될 수 있다. 시스템은 이미지-특징 데이터, 카메라 포즈 데이터 및 관심값 데이터를 지오로케이션 예측 모델에 입력하고, 이미지-특징 데이터, 카메라 포즈 데이터 및 관심값 데이터를 입력하는 것에 응답하여 지오로케이션 데이터를 지오로케이션 예측 모델의 출력으로 얻을 수 있다.

[0027] 일부 구현에서, 지오로케이션 예측 모델은 하나 이상의 분류된 객체 각각과 관련된 단일 임베딩 벡터를 생성할 수 있다. 예를 들어, 단일 임베딩 벡터는 이미지 특징 시퀀스에서 관련 분류된 객체와 관련된 모든 데이터를 인

코딩할 수 있다. 지오로케이션 예측 모델은 단일 임베딩 벡터를 사용하여 관련 분류된 객체와 관련된 실제 위치를 예측할 수 있다.

[0028] 일부 구현에서, 지오로케이션 예측 모델은 프레임 레벨의 위치-특징 추출 모델 및 프레임 레벨의 위치 예측 모델을 포함할 수 있다. 프레임 레벨 위치-특징 추출 모델은 이미지 특징 시퀀스를 나타내는 데이터(예를 들어, 이미지-특징 추출 모델에 의해 출력된 이미지-특징 데이터)를 수신하고, 이미지 특징 시퀀스를 수신하는 것에 응답하여, 하나 이상의 분류된 객체와 관련된 하나 이상의 위치 특징을 포함하는 위치-특징 데이터를 출력하도록 구성될 수 있다. 아래에서 더 설명되는 바와 같이, 본 발명의 양태에 따라, 위치-특징 데이터 내의 하나 이상의 위치 특징은 하나 이상의 분류된 객체에 대한 실제 위치를 예측하는데 사용될 수 있다. 일부 구현에서, 위치-특징 데이터는 위치 특징 시퀀스를 포함할 수 있다. 예를 들어, 위치-특징 데이터내의 하나 이상의 위치 특징은 위치-특징 임베딩 시퀀스로 구성될 수 있다. 위치-특징 임베딩 시퀀스내의 각 위치-특징 임베딩은 이미지 프레임 시퀀스내의 이미지 프레임에 대응할 수 있고, 각각의 위치-특징 임베딩은 대응하는 이미지 프레임에 묘사된 하나 이상의 분류된 객체와 관련된 하나 이상의 위치 특징을 나타낼 수 있다.

[0029] 시스템은 이미지-특징 데이터를 프레임 레벨 위치-특징 추출 모델에 입력하여, 프레임 레벨 위치-특징 추출 모델의 출력으로서 위치-특징 데이터를 획득할 수 있다. 예를 들어, 이미지-특징 데이터는 이미지 특징 시퀀스를 나타내고 이미지 프레임 시퀀스에 대응하는 이미지-특징 임베딩 시퀀스를 포함할 수 있다. 시스템은 이미지 프레임과 관련된 이미지 특징을 나타내는 데이터(예를 들어, 이미지-특징 임베딩 시퀀스내의 이미지-특징 임베딩)를 프레임 레벨 위치-특징 추출 모델에 입력하고, 프레임 레벨 위치-특징 추출 모델의 출력으로서 이미지 프레임과 관련된 (예를 들어, 이미지 프레임에 묘사된 하나 이상의 분류된 객체와 관련된) 하나 이상의 위치 특징을 나타내는 위치-특징 임베딩을 포함하는 위치-특징 데이터를 획득할 수 있다. 이러한 방식으로, 시스템은 이미지-특징 임베딩 시퀀스내의 각 이미지-특징 임베딩을 프레임 레벨 위치-특징 추출 모델에 입력하여, 위치-특징 시퀀스를 나타내고 이미지 프레임 시퀀스에 대응하는 위치-특징 임베딩 시퀀스를 포함하는 위치-특징 데이터를 획득할 수 있다.

[0030] 프레임 레벨 위치 예측 모델은 위치 특징 시퀀스(예를 들어, 프레임 레벨 위치-특징 추출 모델에 의해 출력된 위치-특징 데이터)을 나타내는 데이터를 수신하고, 위치 특징 시퀀스를 수신하는 것에 응답하여 하나 이상의 분류된 객체와 관련된 좌표를 포함하는 좌표 데이터를 출력하도록 구성될 수 있다. 일부 구현에서, 좌표 데이터는 이미지 프레임 시퀀스에 대응하는 좌표 임베딩 시퀀스를 포함할 수 있다. 각 좌표 임베딩은 대응하는 이미지 프레임에 묘사된 하나 이상의 분류된 객체와 관련된 좌표를 나타낼 수 있다. 이미지 프레임에 묘사된 분류된 객체와 관련된 좌표는 이미지 프레임과 관련된 카메라 좌표 공간에서 분류된 객체의 3차원 위치를 나타낼 수 있다.

[0031] 시스템은 위치-특징 데이터를 프레임 레벨 위치 예측 모델에 입력하고, 프레임 레벨 위치 예측 모델의 출력으로서 좌표 데이터를 획득할 수 있다. 예를 들어, 위치-특징 데이터는 위치 특징 시퀀스를 나타내고 이미지 프레임 시퀀스에 대응하는 위치-특징 임베딩 시퀀스를 포함할 수 있다. 시스템은 이미지 프레임과 관련된 위치 특징을 나타내는 데이터(예를 들어, 위치-특징 임베딩 시퀀스내의 위치-특징 임베딩)를 프레임 레벨 위치 예측 모델에 입력하고, 프레임 레벨 위치 예측 모델의 출력으로서 이미지 프레임에 묘사된 하나 이상의 분류된 객체와 관련된 좌표를 나타내는 좌표 임베딩을 포함하는 좌표 데이터를 획득할 수 있다. 이러한 방식으로, 시스템은 위치-특징 임베딩 시퀀스내의 각 위치-특징 임베딩을 프레임 레벨 위치 예측 모델에 입력하고, 좌표 시퀀스를 나타내고 이미지 프레임 시퀀스에 대응하는 좌표 임베딩 시퀀스를 포함하는 좌표 데이터를 획득할 수 있다.

[0032] 지오로케이션 예측 모델은 프레임 레벨 위치 예측 모델 및 카메라 포즈 데이터에 의해 출력된 좌표 데이터에 적어도 부분적으로 기초하여 하나 이상의 분류된 객체와 관련된 예측된 실제 위치를 결정하도록 구성될 수 있다. 일부 구현에서, 지오로케이션 예측 모델은 카메라 좌표 공간의 좌표값 데이터에서 분류된 객체와 관련된 좌표를 실제 좌표(예를 들어, 위도 및 경도)로 변환함으로써 하나 이상의 분류된 객체에 대한 예측된 실제 위치를 결정하도록 구성될 수 있다. 예를 들어, 지오로케이션 예측 모델은 분류된 객체가 이미지 프레임 시퀀에서 복수의 이미지 프레임에 묘사되어 있다고 결정할 수 있다. 지오로케이션 예측 모델은 좌표 데이터에 적어도 부분적으로 기초하여 복수의 이미지 프레임 각각에 대한 분류된 객체와 관련된 좌표를 획득하고, 분류된 객체와 관련된 복수의 좌표를 카메라 포즈 데이터에 적어도 부분적으로 기초하여 실제 좌표(예를 들어, 위도 및 경도)로 변환할 수 있다. 예를 들어, 지오로케이션 예측 모델은 이미지 프레임과 관련된 카메라 좌표 공간내의 분류된 객체의 3차원 위치를 이미지 프레임을 캡처하는데 사용된 카메라의 위치 및/또는 방향에 기초하여 상기 분류된 객체의 실제 좌표로 변환할 수 있다. 이미지 프레임을 캡처한다. 이러한 방식으로, 시스템은 분류된 객체와 관련된 복수의 실제 좌표에 대한 시간적 가중 평균을 결정하여 상기 분류된 객체의 예측된 실제 위치를 결정할 수 있다.

[0033] 일부 구현에서, 지오로케이션 예측 모델은 프레임 레벨 위치 예측 모델에 의해 출력된 좌표 데이터를 검증(확인)하고 그 검증에 기초하여 예측된 실제 좌표를 결정하도록 구성될 수 있다. 예로서, 지오로케이션 예측 모델은 식별된 객체와 관련된 좌표가 정확한지 검증할 수 있다. 다른 예로서, 지오로케이션 예측 모델은 다수의 이미지 프레임에 걸쳐 식별된 객체와 관련된 좌표가 동일한 식별된 객체에 대응하는지 확인할 수 있다.

[0034] 일부 구현에서, 지오로케이션 예측 모델은 예측된 실제 위치가 정확하고 관심있는 분류된 객체에 해당하는지 확인하기 위해, 복수의 손실값 중 하나 이상에 적어도 부분적으로 기초하여 트레이닝될 수 있다. 복수의 손실값은 위치 일관성 손실, 외관 일관성 손실, 에이밍(aiming) 손실 및 FOV(field-of-view) 손실을 포함할 수 있다. 예로서, 시스템은 다수의 이미지 프레임에 걸쳐 식별된 객체와 관련된 좌표들 간의 변화에 적어도 부분적으로 기초하여 위치 일관성 손실을 결정할 수 있다. 시스템은 결정된 위치 일관성 손실을 사용하여, 지오로케이션 예측 모델에 의해 결정된 좌표가 상기 분류된 객체에 대한 다수의 이미지 프레임에 걸쳐 일관되도록 지오로케이션 예측 모델을 트레이닝시킬 수 있다.

[0035] 다른 예로서, 시스템은 (예를 들어, 이미지-특징 추출 모델에 의한 출력된) 이미지-특징 데이터 및 (예를 들어, 객체 분류 모델에 의한 출력된) 관심값 데이터에 적어도 부분적으로 기초하여 외관 일관성 손실을 결정할 수 있다. 특히 시스템은 관심값 데이터에 포함된 공간적 관심값을 갖는 이미지 프레임에 해당하는 이미지 특징들을 가중하여 다수의 이미지 프레임에 대한 외관 특징을 결정할 수 있으며, 시스템은 다수의 이미지 프레임에 걸쳐 상기 결정된 외관 특징 사이의 변화(variance)에 적어도 부분적으로 기초하여 외관 일관성 손실을 결정할 수 있다. 시스템은 결정된 외관 일관성 손실을 사용하여 지오로케이션 예측 모델에 의해 분류된 하나 이상의 객체가 그 객체가 보이는 각 이미지 프레임에서 유사한 시각적 외관을 갖도록 지오로케이션 예측 모델을 트레이닝시킬 수 있다.

[0036] 다른 예로서, 시스템은 (예를 들어, 프레임 레벨 위치 예측 모델에 의해 출력된) 좌표 데이터 및 (예를 들어, 객체 분류 모델에 의해 출력된) 관심값 데이터에 적어도 부분적으로 기초하여 에이밍 손실을 결정할 수 있다. 시스템은 에이밍 손실을 사용하여 지오로케이션 예측 모델을 트레이닝시켜, 이미지 프레임에 묘사된 분류된 객체와 관련된 좌표 데이터의 좌표가 그 분류된 객체와 관련된 공간적 관심이 가장 높은 영역에서 이미지 프레임과 관련된 카메라 좌표 공간에 투사될 수 있도록 할 수 있다.

[0037] 다른 예로서, 시스템은 예측된 실제 좌표가 결정되는 것에 기초하여 이미지 프레임을 캡처하는데 사용된 카메라의 실제 가능한 FOV 내에서 예측된 실제 세계 좌표를 제한하기 위해 FOV 손실을 결정할 수 잇다. 시스템은 예측된 실제 좌표 범위에 의미있는 제한(예를 들어, 합리적인 공간)을 포함하기 위해 결정된 FOV 손실을 사용하여 지오로케이션 예측 모델을 트레이닝할 수 있다.

[0038] 본 명세서에 설명된 시스템 및 방법은 많은 기술적 효과 및 이점을 제공할 수 있다. 예를 들어, 컴퓨팅 시스템은 노이즈 분류를 이용하여 이미지를 나타내는 데이터로부터 위치 정보를 추출할 수 있는 하나 이상의 정보 추출 모델을 포함할 수 있다. 정보 추출 모델(들)은 노이즈 분류로 이미지에 묘사된 하나 이상의 객체의 실제 위치를 예측하기 위해 종단간 트레이닝될 수 있다. 예를 들어 정보 추출 모델(들)은 다양한 유형의 도로 표지판, 집 번호, 회전 제한, 도로명 등의 실제 위치를 예측하는데 사용될 수 있다. 보다 구체적으로, 정보 추출 모델(들)은 객체 분류 모델 및/또는 지오로케이션 예측 모델을 포함할 수 있다. 객체 분류 모델은 지도 신호로서 분류 라벨(예를 들어, 이미지 프레임 시퀀스와 관련된 단일 분류 라벨)이 약한 이미지 데이터를 사용하여 트레이닝될 수 있으며, 지오로케이션 예측 모델은 (예를 들어, 결정된 위치 일관성 손실, 외관 일관성 손실, 에이밍 손실 및/또는 FOV 손실에 기초하여) 비지도 방식으로 트레이닝될 수 있다. 대량의 노이즈 라벨링된 데이터와 지극히 많은 양의 라벨링되지 않은 데이터를 활용하는 기능을 인에이블함으로써, 본 발명은 더 많은 애플리케이션을 위한 모델을 더 빠르고 저렴하게 개발할 수 있다. 또한, 정보 추출 모델에 의해 추출된 위치 정보는 이전에 충분한 지상 실측 위치 데이터가 사용될 수 없었던 많은 새로운 애플리케이션을 개발하는데 사용될 수 있다. 예를 들어, 정보 추출 모델(들)은 차량에 탑재된 컴퓨팅 시스템에 포함되거나 대시캠(dashcam) 애플리케이션의 일부로 포함될 수 있으며, 오프라인 처리를 위해 데이터를 송신할 필요없이 실세계에서 객체들을 검출하는데 사용될 수 있다. 더욱이, 정보 추출 모델(들)의 하나 이상의 컴포넌트(예를 들어, 이미지-특징 추출 모델, 객체 분류 모델, 지오로케이션 예측 모델 등)는 하나 이상의 다른 기계 학습 모델에 통합될 수 있다. 예를 들어, 시공간적 관심를 갖는 객체 분류 모델은 (예를 들어, 비디오 공유 플랫폼에서 폭력적이거나 공격적인 컨텐츠를 분류하기 위해) 노이즈 분류로 다양한 비디오 컨텐츠와 관련된 분류를 결정하는데 사용될 수 있다.

[0039] 예시적인 디바이스 및 시스템

[0040] 도 1a는 본 발명의 예시적인 실시예에 따른 정보 추출을 수행하는 예시적인 지오로케이션 시스템(100)의 블록도

를 도시한다. 특히, 지오로케이션 시스템(100)은 복수의 이미지에 묘사된 하나 이상의 객체에 대한 실제 위치를 예측할 수 있다. 지오로케이션 시스템(100)은 네트워크(180)를 통해 통신적으로 결합된 사용자 컴퓨팅 디바이스(102), 서버 컴퓨팅 시스템(130) 및 트레이닝 컴퓨팅 시스템(150)을 포함하는 컴퓨팅 시스템에 대응할 수 있다.

[0041] 사용자 컴퓨팅 디바이스(102)는 예를 들어 개인 컴퓨팅 디바이스(예를 들어, 랩톱 또는 데스크톱), 모바일 컴퓨팅 디바이스(예를 들어, 스마트 폰 또는 태블릿), 게임 콘솔 또는 컨트롤러, 웨어러블 컴퓨팅 디바이스, 임베디드 컴퓨팅 디바이스 또는 임의의 다른 유형의 컴퓨팅 디바이스와 같은 임의의 유형의 컴퓨팅 디바이스일 수 있다.

[0042] 사용자 컴퓨팅 디바이스(102)는 하나 이상의 프로세서(112) 및 메모리(114)를 포함한다. 하나 이상의 프로세서(112)는 임의의 적절한 처리 디바이스(예를 들어, 프로세서 코어, 마이크로 프로세서, ASIC, FPGA, 컨트롤러, 마이크로 컨트롤러 등)일 수 있고, 하나의 프로세서 또는 동작 가능하게 연결된 복수의 프로세서일 수 있다. 메모리(114)는 RAM, ROM, EEPROM, EPROM, 플래시 메모리 디바이스, 자기 디스크 등 및 이들의 조합과 같은 하나 이상의 비-일시적 컴퓨터 판독 가능 저장 매체를 포함할 수 있다. 메모리(114)는 사용자 컴퓨팅 디바이스(102)가 동작들을 수행하게 하기 위해 프로세서(112)에 의해 실행되는 데이터(116) 및 명령들(118)을 저장할 수 있다.

[0043] 일부 구현에서, 사용자 컴퓨팅 디바이스(102)는 하나 이상의 정보 추출 모델(120)을 저장하거나 포함할 수 있다. 예를 들어, 정보 추출 모델(120)은 신경망(예를 들어, 심층 신경망)과 같은 다양한 기계 학습 모델 또는 비선형 모델 및/또는 선형 모델을 포함하는 다른 유형의 기계 학습 모델일 수 있거나 이를 포함할 수 있다. 신경망에는 피드-포워드 신경망, 순환 신경망(예를 들어, 장단기 메모리 순환 신경망), 컨볼루션 신경망 또는 다른 형태의 신경망이 포함될 수 있다. 예시적인 정보 추출 모델(120)은 도 2-7을 참조하여 논의된다.

[0044] 일부 구현에서, 하나 이상의 정보 추출 모델(120)은 네트워크(180)를 통해 서버 컴퓨팅 시스템(130)으로부터 수신되고, 사용자 컴퓨팅 디바이스 메모리(114)에 저장되고, 그런 다음 하나 이상의 프로세서(112)에 의해 사용되거나 구현될 수 있다. 일부 구현에서, 사용자 컴퓨팅 디바이스(102)는 (예를 들어, 다중 인스턴스에 걸쳐 병렬 정보 추출을 수행하기 위해) 단일 정보 추출 모델(120)의 다중 병렬 인스턴스를 구현할 수 있다.

[0045] 보다 구체적으로, 정보 추출 모델(120)은 이미지 데이터를 수신하고, 이미지 데이터 수신에 응답하여 지오로케이션 데이터를 출력하도록 구성될 수 있다. 지오로케이션 시스템(100)은 정보 추출 모델(120)에 이미지 데이터를 입력하고, 이미지 데이터를 입력하는 것에 응답하여 정보 추출 모델(120)의 출력으로서 지오로케이션 데이터를 획득할 수 있다.

[0046] 추가적으로 또는 대안적으로, 하나 이상의 정보 추출 모델(140)은 클라이언트-서버 관계에 따라 사용자 컴퓨팅 디바이스(102)와 통신하는 서버 컴퓨팅 시스템(130)에 포함되거나 그에 의해 저장 및 구현될 수 있다. 예를 들어, 정보 추출 모델(들)(140)은 웹 서비스(예를 들어, 지오로케이션 정보 추출 서비스)의 일부로서 서버 컴퓨팅 시스템(130)에 의해 구현될 수 있다. 따라서, 하나 이상의 모델(120)이 사용자 컴퓨팅 디바이스(102)에 저장 및 구현될 수 있고 및/또는 하나 이상의 모델(140)이 서버 컴퓨팅 시스템(130)에 저장 및 구현될 수 있다.

[0047] 사용자 컴퓨팅 디바이스(102)는 또한 사용자 입력을 수신하는 하나 이상의 사용자 입력 컴포넌트(122)를 포함할 수 있다. 예를 들어, 사용자 입력 컴포넌트 (122)는 사용자 입력 객체(예를 들어, 손가락 또는 스타일러스)의 터치에 민감한 터치 감지 컴포넌트(예를 들어, 터치 감지 디스플레이 스크린 또는 터치 패드)일 수 있다. 터치 감지 컴포넌트는 가상 키보드를 구현하는 역할을 할 수 있다. 다른 예시적인 사용자 입력 컴포넌트는 마이크로폰, 기존 키보드 또는 사용자가 사용자 입력을 제공할 수 있는 다른 수단을 포함한다.

[0048] 서버 컴퓨팅 시스템(130)은 하나 이상의 프로세서(132) 및 메모리(134)를 포함한다. 하나 이상의 프로세서(132)는 임의의 적절한 처리 디바이스(예를 들어, 프로세서 코어, 마이크로 프로세서, ASIC, FPGA, 컨트롤러, 마이크로 컨트롤러 등) 일 수 있고, 하나의 프로세서 또는 동작 가능하게 연결된 복수의 프로세서일 수 있다. 메모리(134)는 RAM, ROM, EEPROM, EPROM, 플래시 메모리 디바이스, 자기 디스크 등 및 이들의 조합과 같은 하나 이상의 비-일시적 컴퓨터 판독 가능 저장 매체를 포함할 수 있다. 메모리(134)는 서버 컴퓨팅 시스템(130)이 동작들을 수행하게 하기 위해 프로세서(132)에 의해 실행되는 데이터(136) 및 명령들(138)을 저장할 수 있다.

[0049] 일부 구현에서, 서버 컴퓨팅 시스템(130)은 하나 이상의 서버 컴퓨팅 디바이스를 포함하거나 그에 의해 구현된다. 서버 컴퓨팅 시스템(130)이 복수의 서버 컴퓨팅 디바이스를 포함하는 경우, 이러한 서버 컴퓨팅 디바이스는 순차 컴퓨팅 아키텍처, 병렬 컴퓨팅 아키텍처 또는 이들의 일부 조합에 따라 동작할 수 있다.

[0050] 전술한 바와 같이, 서버 컴퓨팅 시스템(130)은 하나 이상의 기계 학습 정보 추출 모델(140)을 저장하거나 이를

포함할 수 있다. 예를 들어, 모델(들)(140)은 다양한 기계 학습 모델일 수 있거나 이를 포함할 수 있다. 예시적인 기계 학습 모델로는 신경망 또는 기타 다층 비선형 모델이 있다. 예시적인 신경망은 피드 포워드 신경망, 심층 신경망, 순환 신경망 및 컨볼루션 신경망이 있다. 예시적인 모델(140)은 도 2-7을 참조하여 논의된다.

[0051] 사용자 컴퓨팅 디바이스(102) 및/또는 서버 컴퓨팅 시스템(130)은 네트워크(180)를 통해 통신적으로 결합된 트레이닝 컴퓨팅 시스템(150)과의 상호 작용을 통해 모델(120 및/또는 140)을 트레이닝시킬 수 있다. 트레이닝 컴퓨팅 시스템 (150)은 서버 컴퓨팅 시스템(130)과 분리될 수 있거나 서버 컴퓨팅 시스템(130)의 일부일 수 있다.

[0052] 트레이닝 컴퓨팅 시스템(150)은 하나 이상의 프로세서(152) 및 메모리(154)를 포함한다. 하나 이상의 프로세서(152)는 임의의 적절한 처리 디바이스(예를 들어, 프로세서 코어, 마이크로 프로세서, ASIC, FPGA, 컨트롤러, 마이크로 컨트롤러 등)일 수 있고, 하나의 프로세서 또는 동작 가능하게 연결된 복수의 프로세서일 수 있다. 메모리(154)는 RAM, ROM, EEPROM, EPROM, 플래시 메모리 디바이스, 자기 디스크 등 및 이들의 조합과 같은 하나 이상의 비-일시적 컴퓨터 판독 가능 저장 매체를 포함할 수 있다. 메모리(154)는 트레이닝 컴퓨팅 시스템(150)이 동작을 수행하게 하기 위해 프로세서(152)에 의해 실행되는 데이터(156) 및 명령들(158)을 저장할 수 있다. 일부 구현에서, 트레이닝 컴퓨팅 시스템(150)은 하나 이상의 서버 컴퓨팅 디바이스를 포함하거나 그에 의해 구현된다.

[0053] 트레이닝 컴퓨팅 시스템(150)은 예를 들어, 오류의 역방향 전파와 같은 다양한 트레이닝 또는 학습 기술을 사용하여 사용자 컴퓨팅 디바이스(102) 및/또는 서버 컴퓨팅 시스템(130)에 저장된 기계 학습된 모델(120 및/또는 140)을 트레이닝하는 모델 트레이너(160)를 포함할 수 있다. 일부 구현에서, 오류의 역방향 전파를 수행하는 것은 시간에 따라 잘린(truncated) 역전파를 수행하는 것을 포함할 수 있다. 모델 트레이너(160)는 트레이닝되는 모델의 일반화 기능을 개선하기 위해 다수의 일반화 기술(예를 들어, 가중치 감쇠, 드롭아웃 등)을 수행할 수 있다.

[0054] 특히, 모델 트레이너(160)는 트레이닝 데이터 세트(162)에 기초하여 정보 추출 모델(120 및/또는 140)을 트레이닝할 수 있다. 예를 들어, 트레이닝 데이터 (162)는 이미지 프레임 시퀀스와 관련된 단일 분류 라벨을 포함하는 이미지 데이터와 같은 약하게 분류된 이미지 데이터를 포함할 수 있다. 모델 트레이너(160)는 약한 분류를 갖는 이미지 데이터를 이용하여 정보 추출 모델(120 및/또는 140)에 포함된 객체 분류 모델을 트레이닝할수 있다. 다른 예로서, 트레이닝 데이터(162)는 이미지 프레임 시퀀스와 관련된 하나 이상의 분류 라벨을 포함할 수 있고, 모델 트레이너(160)는 하나 이상의 분류 라벨을 갖는 이미지 데이터를 사용하여 정보 추출 모델(120 및/또는 140)에 포함 된 객체 분류 모델을 트레이닝할 수 있다. 다른 예로서, 트레이닝 데이터(162)는 정보 추출 모델(120 및/또는 140)에 입력으로 제공되는 데이터 및 그 입력 데이터에 응답하여 정보 추출 모델(120 및/또는 140)의 출력으로 제공되는 데이터를 포함할 수 있다. 모델 트레이너(160)는 입력 데이터와 출력 데이터를 이용하여 정보 추출 모델(120 및/또는 140)에 포함된 지오로케이션 예측 모델을 비지도 방식으로 트레이닝할 수 있다.

[0055] 일부 구현에서, 사용자가 동의를 제공한 경우, 트레이닝 예제가 사용자 컴퓨팅 디바이스(102)에 의해 제공될 수 있다. 따라서, 이러한 구현에서, 사용자 컴퓨팅 디바이스(102)에 제공된 모델(120)은 사용자 컴퓨팅 디바이스(102)로부터 수신된 사용자-특정 데이터에 대해 트레이닝 컴퓨팅 시스템(150)에 의해 트레이닝될 수 있다. 일부 경우, 이 프로세스를 모델 개인화라고 지칭될 수 있다.

[0056] 모델 트레이너(160)는 원하는 기능을 제공하기 위해 사용되는 컴퓨터 로직을 포함한다. 모델 트레이너(160)는 범용 프로세서를 제어하는 하드웨어, 펌웨어 및/또는 소프트웨어로 구현될 수 있다. 예를 들어, 일부 구현에서, 모델 트레이너 (160)는 저장 디바이스에 저장되고, 메모리에 로드되고 하나 이상의 프로세서에 의해 실행되는 프로그램 파일을 포함한다. 다른 구현에서, 모델 트레이너(160)는 RAM 하드 디스크 또는 광학 또는 자기 매체와 같은 유형(tangible)의 컴퓨터 판독 가능 저장 매체에 저장되는 컴퓨터 실행 가능 명령들의 하나 이상의 세트를 포함한다.

[0057] 네트워크(180)는 근거리 네트워크(예를 들어, 인트라넷), 광역 네트워크(예를 들어, 인터넷), 또는 이들의 일부 조합과 같은 임의의 유형의 통신 네트워크 일 수 있으며, 임의의 수의 유선 또는 무선 링크를 포함할 수 있다. 일반적으로 네트워크(180)을 통한 통신은 다양한 통신 프로토콜(예를 들어, TCP/IP, HTTP, SMTP, FTP), 인코딩 또는 포멧(예를 들어, HTML, XML) 및/또는 보호 체계(예를 들어, VPN, 보안 HTTP, SSL)를 사용하여 임의의 유형의 유선 및/또는 무선 연결을 통해 전송될 수 있다.

[0058] 도 1a는 본 발명을 구현하는데 사용될 수 있는 하나의 예시적인 컴퓨팅 시스템을 도시한다. 다른 컴퓨팅 시스템

도 사용될 수 있다. 예를 들어, 일부 구현에서, 사용자 컴퓨팅 디바이스(102)는 모델 트레이너(160) 및 트레이닝 데이터 세트(162)를 포함할 수 있다. 이러한 구현에서, 모델(120)은 사용자 컴퓨팅 디바이스(102)에서 로컬로 트레이닝되고 사용될 수 있다. 이러한 구현 중 일부에서, 사용자 컴퓨팅 디바이스(102)는 사용자-특정 데이터에 기초하여 모델(120)을 개인화하기 위해 모델 트레이너(160)를 구현할 수 있다.

[0059] 도 1b는 본 발명의 예시적인 실시예들에 따른 정보 추출을 수행하는 예시적인 컴퓨팅 디바이스(10)의 블록도를 도시한다. 컴퓨팅 디바이스(10)는 사용자 컴퓨팅 디바이스 또는 서버 컴퓨팅 디바이스일 수 있다.

[0060] 컴퓨팅 디바이스(10)는 다수의 애플리케이션(예를 들어, 애플리케이션 1~N)을 포함한다. 각 애플리케이션은 자체 기계 학습 라이브러리와 기계 학습 모델(들)을 포함한다. 예를 들어, 각 애플리케이션은 기계 학습 모델을 포함할 수 있다. 예시적인 애플리케이션에는 텍스트 메시징 애플리케이션, 이메일 애플리케이션, 구술 애플리케이션, 가상 키보드 애플리케이션, 브라우저 애플리케이션 등이 포함된다.

[0061] 도 1b에 도시된 바와 같이, 각 애플리케이션은 예를 들어 하나 이상의 센서, 컨텍스트 관리자, 디바이스 상태 컴포넌트 및/또는 추가 컴포넌트와 같은 컴퓨팅 디바이스의 다수의 다른 컴포넌트와 통신할 수 있다. 일부 구현에서, 각 애플리케이션은 API(예를 들어, 공공 API)를 사용하여 각 디바이스 컴포넌트와 통신할 수 있다. 일부 구현에서 각 애플리케이션에 의해 사용되는 API는 해당 애플리케이션에 고유(specific, 특정적)하다.

[0062] 도 1c는 본 발명의 예시적인 실시예들에 따른 정보 추출을 수행하는 예시적인 컴퓨팅 디바이스(50)의 블록도를 도시한다. 컴퓨팅 디바이스(50)는 사용자 컴퓨팅 디바이스 또는 서버 컴퓨팅 디바이스일 수 있다.

[0063] 컴퓨팅 디바이스(50)는 다수의 애플리케이션(예를 들어, 애플리케이션 1-N)을 포함한다. 각 애플리케이션은 중앙 인텔리전스 계층(layer)과 통신한다. 예시적인 애플리케이션에는 텍스트 메시징 애플리케이션, 이메일 애플리케이션, 구술 애플리케이션, 가상 키보드 애플리케이션, 브라우저 애플리케이션 등이 포함된다. 일부 구현에서, 각 애플리케이션은 API(예를 들어, 모든 애플리케이션에 걸쳐 공통 API)를 사용하여 중앙 인텔리전스 계층(및 그 안에 저장된 모델(들))과 통신할 수 있다.

[0064] 중앙 인텔리전스 계층은 다수의 기계 학습 모델을 포함한다. 예를 들어, 도 1c에 도시된 바와 같이, 개별 기계 학습 모델(예를 들어, 모델)은 각 애플리케이션에 대해 제공되어 중앙 인텔리전스 계층에 의해 관리될 수 있다. 다른 구현에서, 둘 이상의 애플리케이션은 단일 기계 학습 모델을 공유할 수 있다. 예를 들어 일부 구현에서, 중앙 인텔리전스 계층은 모든 애플리케이션에 대해 단일 기계 학습 모델(예를 들어, 단일 모델)을 제공할 수 있다. 일부 구현에서, 중앙 인텔리전스 계층은 컴퓨팅 디바이스(50)의 운영 체제 내에 포함되거나 그에 의해 구현된다.

[0065] 중앙 인텔리전스 계층은 중앙 디바이스 데이터 계층과 통신할 수 있다. 중앙 디바이스 데이터 계층은 컴퓨팅 디바이스(50)를 위한 중앙 집중식 데이터 저장소 일 수 있다. 도 1c에 도시된 바와 같이, 중앙 디바이스 데이터 계층은 예를 들어 하나 이상의 센서, 컨텍스트 관리자, 디바이스 상태 컴포넌트 및/또는 추가 컴포넌트와 같은 컴퓨팅 디바이스의 다수의 다른 컴포넌트와 통신할 수 있다. 일부 구현에서, 중앙 디바이스 데이터 계층은 API(예를 들어, 개인 API)를 사용하여 각 디바이스 컴포넌트와 통신할 수 있다.

[0066] 도 2는 본 발명의 예시적인 실시예에 따른 예시적인 정보 추출 모델(200)의 블록도를 도시한다. 일부 구현에서, 정보 추출 모델(200)은 예를 들어 이미지 프레임 시퀀스와 같은 복수의 이미지를 기술하는 입력 데이터 세트(204)(예를 들어, 이미지 데이터)를 수신하고, 입력 데이터(204)의 수신 결과로서, 이미지 프레임 시퀀스에 묘사된 하나 이상의 객체에 대한 예측된 실제 위치를 포함하는 출력 데이터(206)(예를 들어, 지오로케이션 데이터)를 제공하도록 트레이닝된다. 일부 구현에서, 입력 데이터(204)는 이미지 프레임 시퀀스내의 하나 이상의 이미지 프레임을 캡처하는데 사용되는 카메라의 실제 위치 및/또는 방향을 나타내는 카메라 포즈 데이터를 포함할 수 있다.

[0067] 도 3은 본 발명의 예시적인 실시예에 따른 예시적인 정보 추출 모델(300)의 블록도를 도시한다. 정보 추출 모델(300)은 정보 추출 모델(300)이 이미지-특징 추출 모델(302), 객체 분류 모델(306) 및 지오로케이션 예측 모델(310)을 추가로 포함하는 것을 제외하고는 도 2의 정보 추출 모델(200)과 유사하다.

[0068] 일부 구현에서, 이미지-특징 추출 모델(302)은 입력 데이터(204) 또는 그 일부(예를 들어, 이미지 프레임 시퀀스를 나타내는 데이터)를 수신하여, 입력 데이터(204)의 수신 결과로서, 입력 데이터(204)의 하나 이상의 이미지 프레임으로부터 추출된 하나 이상의 이미지 특징을 포함하는 이미지-특징 데이터(304)를 제공하도록 트레이닝된다. 일부 구현에서, 이미지-특징 데이터(304)는 하나 이상의 이미지 프레임으로부터 추출된 이미지 특징 시

퀀스를 나타내는 이미지-특징 임베딩 시퀀스를 포함할 수 있다.

[0069] 일부 구현에서, 객체 분류 모델(306)은 이미지-특징 데이터(304)(예를 들어, 이미지 특징 시퀀스를 나타내는 데이터)를 수신하고, 이미지-특징 데이터(304)의 수신의 결과로서, 관심값 데이터(308) 및 분류 데이터(309)를 제공하도록 트레이닝된다. 관심값 데이터(308)는 이미지-특징 데이터(304)의 이미지 특징 시퀀스와 관련된 하나 이상의 시간적 관심값 및 하나 이상의 공간적 관심값을 포함할 수 있다. 분류 데이터(309)는 입력 데이터(204)의 하나 이상의 이미지 프레임에 묘사된 하나 이상의 분류된 객체와 관련된 하나 이상의 분류 라벨(예를 들어, 이미지 프레임에 묘사된 속도 제한 표지판에 대응하는 속도 제한값)을 포함할 수 있다. 일부 구현에서, 관심값 데이터(308)는 하나 이상의 시간적 관심값을 나타내는 시간적 관심 데이터(예를 들어, 도 5에 도시된 바와 같은 시간적 관심 데이터(504)) 및 하나 이상의 공간적 관심값을 나타내는 공간적 관심 데이터(예를 들어, 도 5에 도시된 공간적 관심 데이터(506))를 포함할 수 있다.

[0070] 일부 구현에서, 지오로케이션 예측 모델(310)은 이미지-특징 데이터(304)(예를 들어, 이미지 특성 시퀀스를 나타내는 데이터), 입력 데이터(204) 또는 그의 일부(예를 들어, 카메라 포즈 데이터를 나타내는 데이터) 및 관심값 데이터(308)를 수신하여, 데이터 수신의 결과로서, 입력 데이터(204)에 묘사된(예를 들어, 입력 데이터(204)의 하나 이상의 이미지 프레임에 묘사된) 하나 이상의 객체에 대한 예측된 실제 위치를 포함하는 출력 데이터(206)(예를 들어, 지오로케이션 데이터)를 제공하도록 트레이닝된다. 정보 추출 모델(300)은 분류 데이터(309)에 적어도 부분적으로 기초하여, 객체에 대한 예측된 실제 위치를 그 객체에 대응하는 분류 라벨과 연관시킬 수 있다.

[0071] 도 4는 본 발명의 예시적인 실시예에 따른 예시적인 지오로케이션 예측 모델(400)의 블록도를 도시한다. 지오로케이션 예측 모델(400)은 지오로케이션 예측 모델(400)이 위치 특징 추출 모델(402), 위치 예측 모델(406) 및 좌표 변환 모델(410)을 추가로 포함한다는 점을 제외하고는 도 3의 지오로케이션 예측 모델 (310)과 유사하다.

[0072] 일부 구현에서, 위치 특징 추출 모델(402)은 이미지-특징 데이터(304) 및 관심값 데이터(308)를 수신하여, 데이터 수신의 결과로서, 입력 데이터(204)에 묘사된 하나 이상의 분류된 객체와 관련된 하나 이상의 위치 특징을 포함하는 위치-특징 데이터(404)를 제공하도록 트레이닝된다. 일부 구현에서, 위치-특징 데이터는 입력 데이터(204) 내의 이미지 프레임 시퀀스에 대응하는 위치-특징 임베딩 시퀀스를 포함할 수 있다. 위치-특징 임베딩 시퀀스내의 각 위치-특징 임베딩은 대응하는 이미지 프레임에 묘사된 하나 이상의 분류된 객체와 관련된 위치 특성을 나타낼 수 있다.

[0073] 일부 구현에서, 위치 예측 모델(406)은 위치-특징 데이터(404)를 수신하여, 위치-특징 데이터(404)의 수신의 결과로서, 입력 데이터(204)에 묘사된 하나 이상의 분류된 객체와 관련된 좌표를 포함하는 좌표 데이터(408)를 제공하도록 트레이닝된다. 일부 구현에서, 좌표 데이터(408)는 입력 데이터(204) 내의 이미지 프레임 시퀀스에 대응하는 좌표 임베딩 시퀀스를 포함할 수 있다. 좌표 임베딩 시퀀스의 각 좌표 임베딩은 대응하는 이미지 프레임에 묘사된 하나 이상의 분류된 객체와 관련된 좌표를 나타낼 수 있다. 일부 구현에서, 좌표 데이터(408)는 분류된 객체를 묘사하는 이미지 프레임과 관련된 카메라 좌표 공간내의 분류된 객체의 3차원 위치를 나타내는 분류된 객체와 관련된 좌표를 포함할 수 있다.

[0074] 일부 구현에서, 좌표 변환 모델(410)은 좌표 데이터(408) 및 입력 데이터 (204)의 적어도 일부(예를 들어, 카메라 포즈 데이터)를 수신하여, 데이터 수신의 결과로서 출력 데이터(206)(예를 들어, 지오로케이션 데이터)를 제공하도록 트레이닝된다. 특히, 좌표 변환 모델(410)은 카메라 좌표 공간내의 분류된 객체와 관련된 좌표를 실제 좌표(예를 들어, 위도 및 경도 값)로 변환할 수 있다.

[0075] 일부 구현에서, 지오로케이션 예측 모델(400)은 예측된 실제 위치가 정확하고 관심있는 분류된 객체에 대응하는지 확인하기 위해 복수의 손실값 중 하나 이상에 적어도 부분적으로 기초하여 트레이닝될 수 있다. 예를 들어, 지오로케이션 시스템(100)은 다수의 이미지 프레임에 걸쳐 식별된 객체와 관련된 좌표 간의 변화에 적어도 부분적으로 기초하여 위치 일관성 손실을 결정할 수 있다. 지오로케이션 시스템(100)은 결정된 위치 일관성 손실을 사용하여, 지오로케이션 예측 모델에 의해 결정된 좌표들이 분류된 객체에 대한 다수의 이미지 프레임에 걸쳐 일관되도록 지오로케이션 예측 모델(400)을 트레이닝시킬 수 있다. 다른 예로서, 지오로케이션 시스템(100)은 이미지-특징 데이터(304) 및 관심값 데이터(308)에 적어도 부분적으로 기초하여 외관 일관성 손실을 결정할 수 있다. 특히, 지오로케이션 시스템 (100)은 다수의 이미지 프레임에 대한 외관 특징을 결정하기 위해 관심값 데이터(308)에 포함된 공간적 관심값를 갖는 이미지 프레임에 대응하는 이미지 특징을 가중할 수 있고, 지오로케이션 시스템(100)은 다수의 이미지 프레임에 걸쳐 그 결정된 외관 특징들 사이의 변화에 적어도 부분적으로 기초하여 외관 일관성 손실을 결정할 수 있다. 지오로케이션 시스템(100)은 지오로케이션 예측 모델에 의해 분류

된 하나 이상의 객체가 객체가 그 객체가 보이는 각각의 이미지 프레임에서 유사한 시각적 외관을 갖도록 상기 결정된 외관 일관성 손실을 사용하여 지오로케이션 예측 모델(400)을 트레이닝할 수 있다. 다른 예로서, 지오로케이션 시스템(100)은 좌표 데이터(408) 및 관심값 데이터(308)에 적어도 부분적으로 기초하여 에이밍 손실을 결정할 수 있다. 지오로케이션 시스템(100)은 이미지 프레임에 묘사된 분류된 객체와 관련된 좌표 데이터(408)의 좌표들이 그 분류된 객체와 관련된 공간적 관심이 가장 높은 영역의 이미지 프레임과 관련된 카메라 좌표 공간에 투사되도록 에이밍 손실을 사용하여 지오로케이션 예측 모델(400)을 트레이닝시킨다. 다른 예로서, 지오로케이션 시스템(100)은 예측된 실제 좌표가 결정되는 것에 기초하여 이미지 프레임을 캡처하는데 사용된 카메라의 실제 가능한 FOV 내에서 예측된 실제 좌표를 제한하기 위해 시야(FOV) 손실을 결정할 수 있다. 지오로케이션 시스템 (100)은 예측된 실제 좌표의 범위에 의미있는 한계(예를 들어, 합리적인 공간)를 포함하기 위해 상기 결정된 FOV 손실을 사용하여 지오로케이션 예측 모델(400)을 트레이닝할 수 있다.

[0076] 도 5는 본 발명의 예시적인 실시예에 따른 예시적인 객체 분류 모델(500)의 블록도를 도시한다. 객체 분류 모델(500)은 객체 분류 모델(500)이 관심값 데이터(308)에 추가하여 분류 데이터(309)를 출력하는 것을 제외하고는 도 3의 객체 분류 모델(306)과 유사하다.

[0077] 일부 구현에서, 객체 분류 모델(500)은 이미지 특징 시퀀스(예를 들어, 이미지-특징 데이터(304))를 나타내는 데이터를 수신하여, 이미지-특징 데이터(304)의 수신의 결과로서, 분류 데이터(309), 시간적 관심 데이터(504) 및 공간적 관심 데이터(506)를 제공하도록 트레이닝된다. 분류 데이터(309)는 이미지 프레임 시퀀스에 묘사된 하나 이상의 객체와 관련된 분류를 포함할 수 있다. 시간적 관심 데이터(504)는 이미지 특징 시퀀스와 관련된 하나 이상의 시간적 관심값을 포함할 수 있고, 공간적 관심 데이터(506)는 이미지 특징 시퀀스와 관련된 하나 이상의 공간적 관심값를 포함할 수 있다. 일부 구현에서, 객체 분류 모델(500)은 분류 데이터(309)에 적어도 부분적으로 기초하여 트레이닝될 수 있다. 예를 들어, 지오로케이션 시스템(100)은 분류된 데이터 내의 하나 이상의 분류 라벨 및 입력 데이터(204) 내의 이미지 프레임 시퀀스와 관련된 분류에 적어도 부분적으로 기초하여 소프트 맥스 교차 엔트로피 손실을 결정할 수 있다. 지오로케이션 시스템(100)은 결정된 소프트 맥스 교차 엔트로피 손실에 적어도 부분적으로 기초하여 객체 분류 모델(500)을 트레이닝할 수 있다.

[0078] 일부 구현에서, 객체 분류 모델(500)은 시공간적 관심 메커니즘 계층(510), 복수의 LSTM 블록을 포함하는 장단기 메모리(LSTM) 계층(512), 및 완전 연결(fully connected : FC) 계층(514)을 포함할 수 있다. 시공간적 관심 메커니즘 계층(510)은 이미지-특징 데이터(304)에 적어도 부분적으로 기초하여 시간적 관심 데이터(504) 및 공간적 관심 데이터(506)를 결정할 수 있다. LSTM 계층(512)의 각 LSTM 블록은 이미지-특징 데이터(304)에 적어도 부분적으로 기초하여 프레임 당 임베딩을 결정하고, 다중 이미지 프레임에 걸쳐 지속되는 하나 이상의 객체를 결정하기 위해 프레임 당 임베딩을 FC 계층(514)에 제공할 수 있다. 객체 분류 모델(500)은 분류 데이터(309)를 결정하기 위해 시간적 관심 데이터(504)에 기초하여 프레임 당 임베딩을 가중할 수 있다.

[0079] 도 6은 본 발명의 예시적인 실시예에 따른 예시적인 위치 특징 추출 모델 (600)의 블록도를 도시한다. 위치 특징 추출 모델(600)은 위치 특징 추출 모델 (600)이 단일 이미지 프레임에 대응하는 하나 이상의 이미지 특징(예를 들어, 이미지-특징 데이터(304))을 나타내는 데이터를 수신하도록 트레이닝된다는 점을 제외하면 도 4의 위치 특징 추출 모델(402)과 유사하다. 하나 이상의 이미지 특징 및 관심값 데이터(308)의 수신의 결과로서, 위치 특징 추출 모델(600)은 단일 이미지 프레임에 대응하는 하나 이상의 위치 특징을 포함하는 위치-특징 데이터(404)를 제공한다. 일부 구현에서, 지오로케이션 시스템(100)은 입력 데이터(204)의 각 이미지 프레임에 대해, 이미지 프레임에 대응하는 하나 이상의 이미지 특징을 나타내는 데이터를 순차적으로 입력할 수 있다. 일부 구현에서, 정보 추출 모델(300)은 복수의 위치 특징 추출 모델(600)을 포함할 수 있고, 지오로케이션 시스템(100)은 하나 이상의 이미지 특징을 나타내는 데이터를 병렬로 입력할 수 있다. 예를 들어, 정보 추출 모델(300)이 제1 및 제2 위치 특징 추출 모델(600)을 포함한다면, 지오로케이션 시스템(100)은 제1 이미지 프레임에 대응하는 하나 이상의 이미지 특징을 나타내는 데이터를 제1 위치 특징 추출 모델(600)에 그리고 제2 이미지 프레임에 대응하는 하나 이상의 이미지 특징을 나타내는 데이터를 제2 위치 특징 추출 모델(600)에 동시에 입력할 수 있다. 이러한 방식으로, 위치 특징 추출 모델(600)은 입력 데이터(204)의 이미지 프레임 시퀀스에 대응하는 위치 특징 시퀀스(예를 들어, 위치-특징 데이터(404))을 제공할 수 있다.

[0080] 도 7은 본 발명의 예시적인 실시예에 따른 예시적인 위치 예측 모델(700)의 블록도를 도시한다. 위치 예측 모델(700)은 위치 예측 모델(700)이 복수의 LSTM 블록을 포함하는 장단기 메모리(LSTM) 계층(712) 및 완전 연결(FC) 계층(714)을 포함한다는 점을 제외하면 도 4의 위치 예측 모델(406)과 유사하다. 위치 예측 모델(700)은 입력 데이터(204)(예를 들어, 위치-특징 데이터(404))의 이미지 프레임 시퀀스에 대응하는 위치 특징 시퀀스를 나타내는 데이터를 수신하고, 위치 특징 시퀀스를 수신한 결과로서 위치 예측 모델(700)이 이미지 프레임 시퀀스에

묘사된 분류된 객체에 대한 이미지 프레임 시퀀스에 대응하는 좌표 시퀀스를 나타내는 데이터를 제공하도록 트레이닝된다. 좌표 시퀀스는 예를 들어 분류된 객체를 묘사하는 각 이미지 프레임내의 분류된 객체와 관련된 좌표를 포함할 수 있다. 예를 들어, 위치 예측 모델(700)은 위치-특징 임베딩 시퀀스를 포함하는 위치-특징 데이터(404)를 수신할 수 있으며, 각 위치-특징 임베딩은 이미지 프레임 시퀀스의 이미지 프레임에 대응하는 하나 이상의 위치 특성을 나타낸다. 위치 예측 모델(700)은 LSTM 계층(712)의 대응하는 LSTM 블록에 각각의 위치-특징 임베딩을 제공할 수 있다. 각 LSTM 블록의 출력은 대응하는 이미지 프레임에서 객체의 예측된 위치를 나타낼 수 있다. 이러한 방식으로, LSTM 계층(712)은 객체에 대한 예측된 위치 시퀀스를 출력할 수 있는데, 그 예측된 위치 시퀀스는 객체를 묘사하는 이미지 프레임 시퀀스에서 각 이미지 프레임내의 객체의 예측된 위치에 대응한다. LSTM 계층(712)의 출력은 객체에 대한 좌표 시퀀스를 포함하는 좌표 데이터(408)를 결정하기 위해 FC 계층(714)에 제공될 수 있다.

[0081] 일부 구현에서, 지오로케이션 시스템(100)은 위치 예측 모델(700)을 사용하여, 입력 데이터(204)의 이미지 프레임 시퀀스에 묘사된 복수의 분류된 객체에 대한 좌표 시퀀스를 순차적으로 결정할 수 있다. 예를 들어, 위치 예측 모델(700)의 각각의 반복은 이미지 프레임 시퀀스에 묘사된 상이한 객체와 관련된 좌표 시퀀스를 출력할 수 있다. 일부 구현에서, 정보 추출 모델(300)은 복수의 위치 예측 모델(700)을 포함할 수 있고, 지오로케이션 시스템(100)은 위치-특징 데이터(404)를 복수의 위치 예측 모델(700) 각각에 병렬로 입력할 수 있다. 예를 들어, 정보 추출 모델(300)이 제1 및 제2 위치 예측 모델(700)을 포함하는 경우, 지오로케이션 시스템(100)은 위치-특징 데이터(404)를 제1 및 제2 위치 예측 모델(700)에 동시에 입력하여, 제1 위치 예측 모델(700)의 출력으로서 제1 분류된 객체와 관련된 제1 좌표 시퀀스 및 제2 위치 예측 모델(700)의 출력으로서 제2 분류된 객체와 관련된 제2 좌표 시퀀스를 획득할 수 있다.

[0082] 예시적인 방법

[0083] 도 8은 본 발명의 예시적인 실시예에 따른 정보 추출을 수행하기 위한 예시적인 방법의 흐름도를 도시한다. 도 8은 예시 및 논의를 위해 특정 순서로 수행되는 단계들을 도시하지만, 본 발명의 방법은 특별히 예시된 순서 또는 배열로 제한되지 않는다. 방법(800)의 다양한 단계는 본 발명의 범위를 벗어나지 않고 다양한 방식으로 생략, 재배열, 결합 및/또는 적응될 수 있다.

[0084] 단계(802)에서, 컴퓨팅 시스템은 이미지 시퀀스를 나타내는 데이터를 얻을 수 있다. 예를 들어, 지오로케이션 시스템(100)은 이미지 시퀀스를 나타내는 데이터를 포함하는 입력 데이터(204)를 획득할 수 있다. 지오로케이션 시스템(100)은 이미지 시퀀스로부터 위치 정보를 추출하도록 트레이닝된 기계 학습된 정보 추출 모델(120/140)에 이미지 시퀀스를 입력할 수 있다. 일부 구현에서, 이미지 시퀀스는 이미지 시퀀스의 다수의 이미지에 걸쳐 복수의 객체를 묘사할 수 있고, 정보 추출 모델(120/140)의 출력은 이미지 시퀀스에 묘사된 복수의 객체와 관련된 실제 위치를 나타내는 데이터를 포함할 수 있다.

[0085] 단계(804)에서, 컴퓨팅 시스템은 이미지 시퀀스로부터 추출된 이미지 특징 시퀀스에 적어도 부분적으로 기초하여 이미지 시퀀스와 관련된 분류 라벨들 및 관심값들을 결정할 수 있다. 예를 들어, 지오로케이션 시스템(100)은 (예를 들어, 이미지-특징 추출 모델(302)에 의해) 이미지 시퀀스로부터 추출된 이미지 특징 시퀀스(예를 들어, 이미지-특징 데이터(304))를 나타내는 데이터에 적어도 부분적으로 기초하여, 분류 데이터(309), 시간적 관심값을 포함하는 시간적 관심 데이터 (504), 및 이미지 시퀀스와 관련된 공간적 관심값를 포함하는 공간적 관심 데이터(506)를 결정할 수 있다. 특히, 지오로케이션 시스템(100)은 이미지 특징 시퀀스를 약하게 지도된 객체 분류 모델(306)에 입력하고, 그 이미지 특징 시퀀스를 입력하는 것에 응답하여 객체 분류 모델(306)의 출력으로서, 분류 데이터(309), 시간적 관심 데이터(504) 및 공간적 관심 데이터(506)를 획득할 수 있다. 지오로케이션 시스템(100)은 이미지 특징 시퀀스, 시간적 관심 데이터(504) 및 공간적 관심 데이터(506)에 적어도 부분적으로 기초하여 객체와 관련된 실제 위치를 예측할 수 있다.

[0086] 단계(806)에서, 컴퓨팅 시스템은 이미지 특징 시퀀스 및 관심값에 적어도 부분적으로 기초하여 위치 특징 시퀀스를 결정할 수 있다. 예를 들어, 지오로케이션 시스템(100)은 이미지 특징 시퀀스, 시간적 관심 데이터(504) 및 공간적 관심 데이터(506)를 나타내는 데이터를 프레임 레벨 위치-특징 추출 모델(600)에 입력하고, 이미지 특징 시퀀스, 시간적 관심 데이터(504) 및 공간적 관심 데이터(506)를 입력하는 것에 응답하여 프레임 레벨 위치-특징 추출 모델(600)의 출력으로서, 객체와 관련된 하나 이상의 위치 특징을 포함하는 위치 특징 시퀀스를 나타내는 위치-특징 데이터(404)를 획득할 수 있다.

[0087] 단계(808)에서, 컴퓨팅 시스템은 위치 특징 시퀀스 및 관심값들에 적어도 부분적으로 기초하여 이미지 시퀀스에 묘사된 하나 이상의 객체와 관련된 좌표를 결정할 수 있다. 예를 들어, 지오로케이션 시스템(100)은 위치-특징

데이터(404)를 프레임 레벨 위치 예측 모델(406)에 입력하고, 위치-특징 데이터(404)의 입력에 응답하여 프레임 레벨 위치 예측 모델(406)의 출력으로서, 객체와 관련된 카메라 좌표 공간의 좌표를 나타내는 좌표 데이터(408)를 획득할 수 있다. 지오로케이션 시스템(100)은 좌표 데이터(408) 및 입력 데이터(204)의 객체와 관련된 카메라 포즈 데이터에 적어도 부분적으로 기초하여 객체와 관련된 실제 좌표를 결정할 수 있다.

[0088] 단계(810)에서, 컴퓨팅 시스템은 결정된 좌표에 적어도 부분적으로 기초하여 하나 이상의 객체와 관련된 실제 위치를 예측할 수 있다. 예를 들어, 지오로케이션 시스템(100)은 입력 데이터(204)의 입력에 응답하여 정보 추출 모델(120/140)의 출력으로서, 이미지 시퀀스에 묘사된 객체와 관련된 실제 위치를 나타내는 출력 데이터(206)를 획득할 수 있다. 지오로케이션 시스템(100)은 분류 데이터(309)에 적어도 부분적으로 기초하여, 객체에 대한 예측된 실제 위치를 그 객체에 대응하는 분류 라벨과 연관시킬 수 있다.

[0089] 도 9는 본 발명의 예시적인 실시예에 따른 정보 추출 모델을 트레이닝하기 위한 예시적인 방법의 흐름도를 도시한다. 도 9는 예시 및 논의의 목적으로 특정 순서로 수행되는 단계를 도시하지만, 본 발명의 방법은 특별히 예시된 순서 또는 배열로 제한되지 않는다. 방법(900)의 다양한 단계들은 본 발명의 범위를 벗어나지 않고 다양한 방식으로 생략, 재배열, 결합 및/또는 적응될 수 있다.

[0090] 단계(902)에서, 컴퓨팅 시스템(예를 들어, 트레이닝 컴퓨팅 시스템(150) 또는 지오로케이션 시스템(100)의 다른 부분)은 노이즈 분류를 이용하여 이미지 시퀀스로부터 추출된 이미지 특징 시퀀스를 나타내는 데이터를 획득할 수 있다. 예를 들어, 지오로케이션 시스템(100)은 이미지 시퀀스와 관련된 단일 분류 라벨을 갖는 이미지 시퀀스로부터 추출된 이미지 특징 시퀀스(예를 들어, 이미지-특징 데이터(304))를 나타내는 이미지 데이터를 획득할 수 있다. 지오로케이션 시스템(100)은 이미지 데이터를 이미지-특징 추출 모델(302)에 입력하고, 이미지 데이터를 입력하는 것에 응답하여 이미지-특징 추출 모델(302)의 출력으로서 이미지-특징 데이터(304)를 획득할 수 있다.

[0091] 단계(904)에서, 컴퓨팅 시스템은 이미지 특징 시퀀스에 적어도 부분적으로 기초하여 이미지 시퀀스에 묘사된 하나 이상의 객체와 관련된 분류를 결정할 수 있다. 예를 들어, 지오로케이션 시스템(100)은 이미지 특징 데이터(304)를 약하게 지도된 객체 분류 모델(306)에 입력하고, 이미지 특징 시퀀스를 입력하는 것에 응답하여 객체 분류 모델(306)의 출력으로서, 이미지 프레임 시퀀스에 묘사된 객체와 관련된 분류를 나타내는 데이터(예를 들어, 분류 데이터(309))를 획득할 수 있다.

[0092] 단계(906)에서, 컴퓨팅 시스템은 결정된 분류와 관련된 손실을 결정할 수 있다. 예를 들어, 지오로케이션 시스템(100)은 이미지 시퀀스와 관련된 노이즈 분류에 적어도 부분적으로 기초하여 객체 분류 모델(306)에 의해 출력된 분류 데이터(309)와 관련된 손실을 결정할 수 있다.

[0093] 단계(908)에서, 컴퓨팅 시스템은 결정된 분류와 관련된 손실에 적어도 부분적으로 기초하여 객체 분류 모델을 트레이닝시킬 수 있다. 예를 들어, 지오로케이션 시스템(100)은 결정된 손실에 적어도 부분적으로 기초하여 객체 분류 모델(306)을 트레이닝시킬 수 있다.

[0094] 도 10은 본 발명의 내용의 예시적인 실시예에 따른 정보 추출 모델을 트레이닝하기 위한 예시적인 방법의 흐름도를 도시한다. 도 10은 예시 및 논의를 위해 특정 순서로 수행되는 단계들을 도시하지만, 본 발명의 방법은 특별히 예시된 순서 또는 배열로 제한되지 않는다. 방법(1000)의 다양한 단계들은 본 발명의 범위를 벗어나지 않고 다양한 방식으로 생략, 재배열, 결합 및/또는 적응될 수 있다.

[0095] 단계(1002)에서, 컴퓨팅 시스템(예를 들어, 트레이닝 컴퓨팅 시스템(150) 또는 지오로케이션 시스템(100)의 다른 부분)은 노이즈 분류를 이용하여 이미지 시퀀스로부터 추출된 이미지 특징 시퀀스를 나타내는 데이터 및 이미지 시퀀스와 관련된 관심값를 나타내는 데이터를 얻을 수 있다. 예를 들어, 지오로케이션 시스템(100)은 이미지 시퀀스(예를 들어, 입력 데이터(204))로부터 추출된 이미지 특징 시퀀스(예를 들어, 이미지-특징 데이터(304))을 나타내는 데이터를 획득할 수 있다. 지오로케이션 시스템(100)은 입력 데이터(204)를 이미지-특징 추출 모델(302)에 입력하고, 입력 데이터(204) 입력에 응답하여 이미지-특징 추출 모델(302)의 출력으로서 이미지-특징 데이터(304)를 획득할 수 있다.

[0096] 단계(1004)에서, 컴퓨팅 시스템은 위치 특징 시퀀스 및 관심값들에 적어도 부분적으로 기초하여 이미지 시퀀스에 묘사된 하나 이상의 객체와 관련된 실제 위치를 예측할 수 있다. 예를 들어, 지오로케이션 시스템(100)은 입력 데이터(204) 입력에 응답하여 정보 추출 모델(120/140)의 출력으로서 출력 데이터(206)를 획득할 수 있다. 출력 데이터(206)는 이미지 시퀀스에 묘사된 객체와 관련된 실제 위치를 나타낼 수 있다. 특히, 지오로케이션 시스템(100)은 위치 특징 추출 모델(402)에 이미지-특징 데이터(304)와 관심값 데이터(308)를 입력할 수 있고,

결과로서, 입력 데이터(204)에 묘사된 하나 이상의 분류된 객체와 관련된 하나 이상의 위치 특징을 포함하는 위치-특징 데이터(404)를 획득할 수 있다. 지오로케이션 시스템(100)은 위치-특징 데이터(404)를 위치 예측 모델(406)에 입력하고, 결과로서, 입력 데이터(204)에 묘사된 하나 이상의 분류된 객체와 관련된 좌표를 포함하는 좌표 데이터(408)를 획득할 수 있다. 지오로케이션 시스템(100)은 좌표 데이터(408) 및 입력 데이터(204)의 적어도 일부(예를 들어, 카메라 포즈 데이터)를 좌표 변환 모델(410)에 입력하고, 결과로서, 입력 데이터(204)에 묘사된 하나 이상의 객체(예를 들어, 거리 표지판)에 대한 예측된 실제 위치(예를 들어, 위도 및 경도)를 포함하는 출력 데이터(206)(예를 들어, 지오로케이션 데이터)를 획득할 수 잇다.

[0097] 단계(1006)에서, 컴퓨팅 시스템은 하나 이상의 객체와 관련된 예측된 실제 위치에 적어도 부분적으로 기초하여 위치 일관성 손실을 결정할 수 있다. 예를 들어, 지오로케이션 시스템(100)은 객체를 묘사하는 이미지 시퀀스에서 다수의 이미지에 걸쳐 객체와 관련된 좌표들 사이의 변화에 적어도 부분적으로 기초하여 위치 일관성 손실을 결정할 수 있다.

[0098] 단계(1008)에서, 컴퓨팅 시스템은 하나 이상의 객체와 관련된 예측된 실제 위치에 적어도 부분적으로 기초하여 외관 일관성 손실을 결정할 수 있다. 예를 들어, 지오로케이션 시스템(100)은 객체를 묘사하는 이미지 시퀀스에서 다수의 이미지에 걸쳐 결정된 외관 특징들 사이의 변화에 적어도 부분적으로 기초하여 외관 일관성 손실을 결정할 수 있다.

[0099] 단계(1010)에서, 컴퓨팅 시스템은 하나 이상의 객체와 관련된 예측된 실제 위치에 적어도 부분적으로 기초하여 에이밍 손실을 결정할 수 있다. 예를 들어, 지오로케이션 시스템(100)은 객체와 관련된 카메라 좌표 공간의 좌표 및 객체를 묘사하는 이미지 시퀀스에서 다수의 이미지에 걸쳐 객체와 관련된 공간적 관심에 적어도 부분적으로 기초하여 에이밍 손실을 결정할 수 있다.

[0100] 단계(1012)에서, 컴퓨팅 시스템은 하나 이상의 객체와 관련된 예측된 실제 위치에 적어도 부분적으로 기초하여 시야 손실을 결정할 수 있다. 예를 들어, 지오로케이션 시스템(100)은 객체와 관련된 실제 좌표와 객체를 묘사하는 이미지 시퀀스를 캡처하는데 사용된 카메라와 관련된 시야에 적어도 부분적으로 기초하여 시야 손실을 결정할 수 있다.

[0101] 단계(1014)에서, 컴퓨팅 시스템은 결정된 손실에 적어도 부분적으로 기초하여 위치 예측 모델을 트레이닝할 수 있다. 예를 들어, 지오로케이션 시스템(100)은 위치 일관성 손실, 외관 일관성 손실, 에이밍 손실 및/또는 시야 손실에 적어도 부분적으로 기초하여 위치 예측 모델(406)을 트레이닝할 수 있다.

[0102] 추가적인 개시

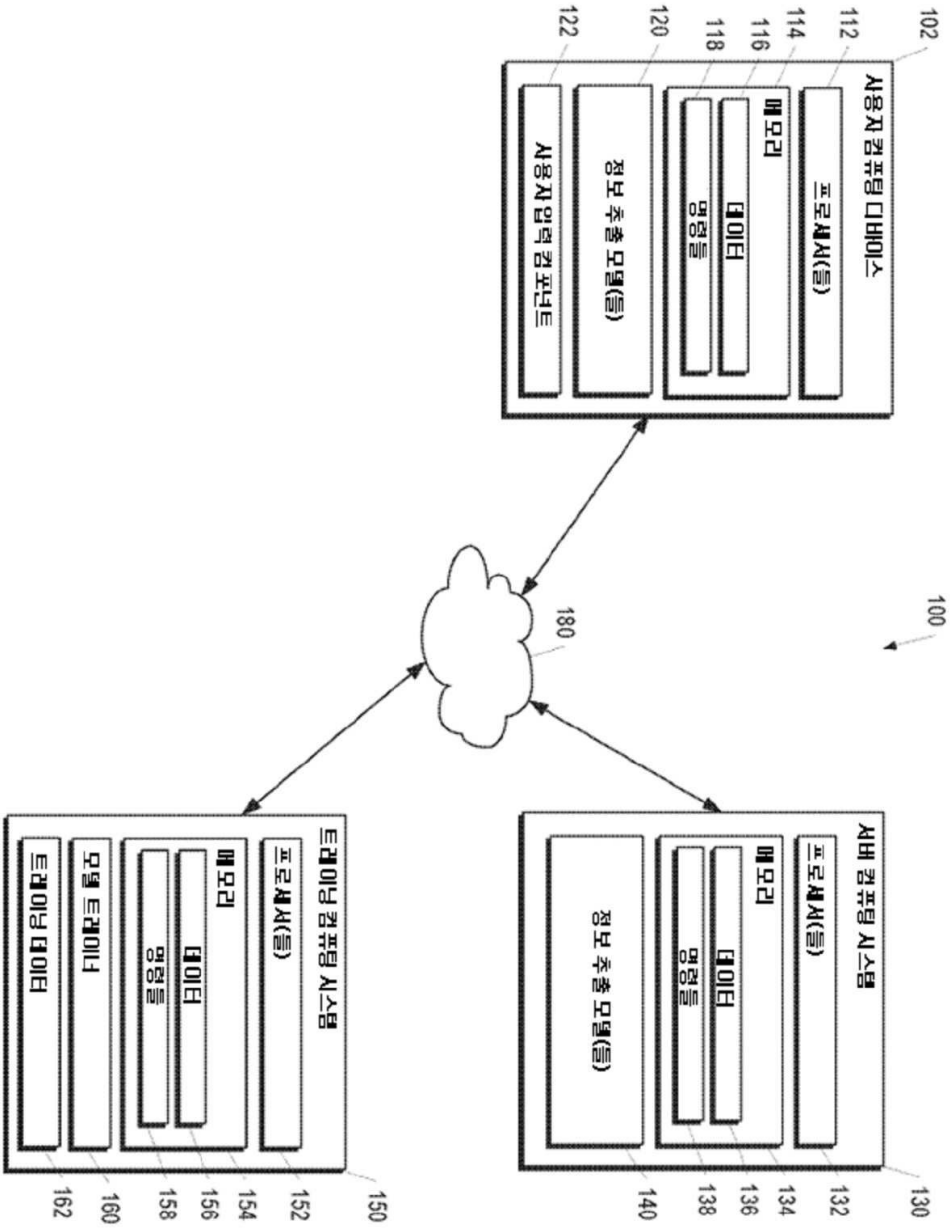
[0103] 본 명세서에서 논의된 기술은 서버, 데이터베이스, 소프트웨어 애플리케이션 및 기타 컴퓨터 기반 시스템뿐만 아니라 이러한 시스템으로(부터) 전송되는 동작 및 정보를 참조한다. 컴퓨터 기반 시스템의 고유한 유연성을 통해 컴포넌트 간의 작업 및 기능의 매우 다양한 구성, 조합 및 분할을 기능하게 한다. 예를 들어, 본 명세서에 설명된 프로세스는 단일 디바이스 또는 컴포넌트 또는 조합하여 작동하는 다수의 디바이스 또는 컴포넌트를 사용하여 구현될 수 있다. 데이터베이스와 애플리케이션은 단일 시스템에서 구현되거나 다수의 시스템에 걸쳐 분산될 수 있다. 분산된 컴포넌트는 순차적으로 또는 병렬로 작동할 수 있다.

[0104] 본 발명은 다양한 특정 실시예에 대하여 상세하게 설명하였으나, 각각의 예는 본 발명을 한정하는 것이 아니라 설명을 위한 것이다. 당업자는 전술한 내용을 이해하면 이러한 실시예에 대한 변경, 변형 및 등가물을 쉽게 생성할 수 있다. 따라서, 본 발명의 내용은 당업자에게 쉽게 명백할 바와 같이 현재의 주제에 대한 이러한 수정, 변경 및/또는 추가의 포함을 배제하지 않는다. 예를 들어, 일 실시예의 일부로서 예시되거나 설명된 특징은 또 다른 실시예와 함께 사용되어 또 다른 실시예를 생성할 수 있다. 따라서, 본 발명의 내용은 이러한 변경, 변형 및 등가물을 포함하는 것으로 의도된다.

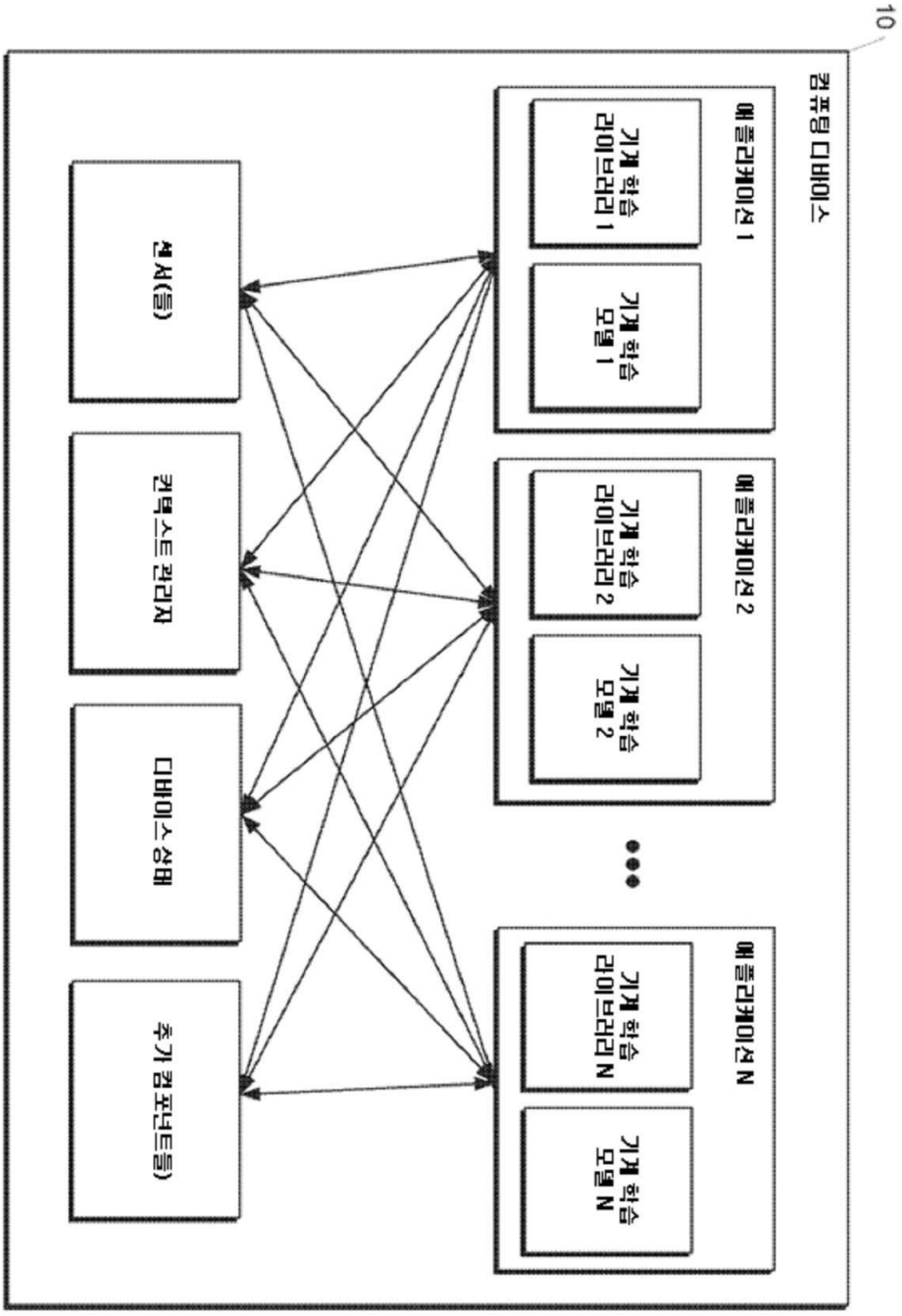
[0105] 특히, 도 8 내지 10은 각각 예시 및 논의의 목적을 위해 특정 순서로 수행되는 단계를 도시하지만, 본 개시 내용의 방법은 특별히 예시된 순서 또는 배열로 제한되지 않는다. 방법(800, 900 및 1000)의 다양한 단계는 본 발명의 범위를 벗어나지 않고 다양한 방식으로 생략, 재배열, 결합 및/또는 적응될 수 있다.

도면

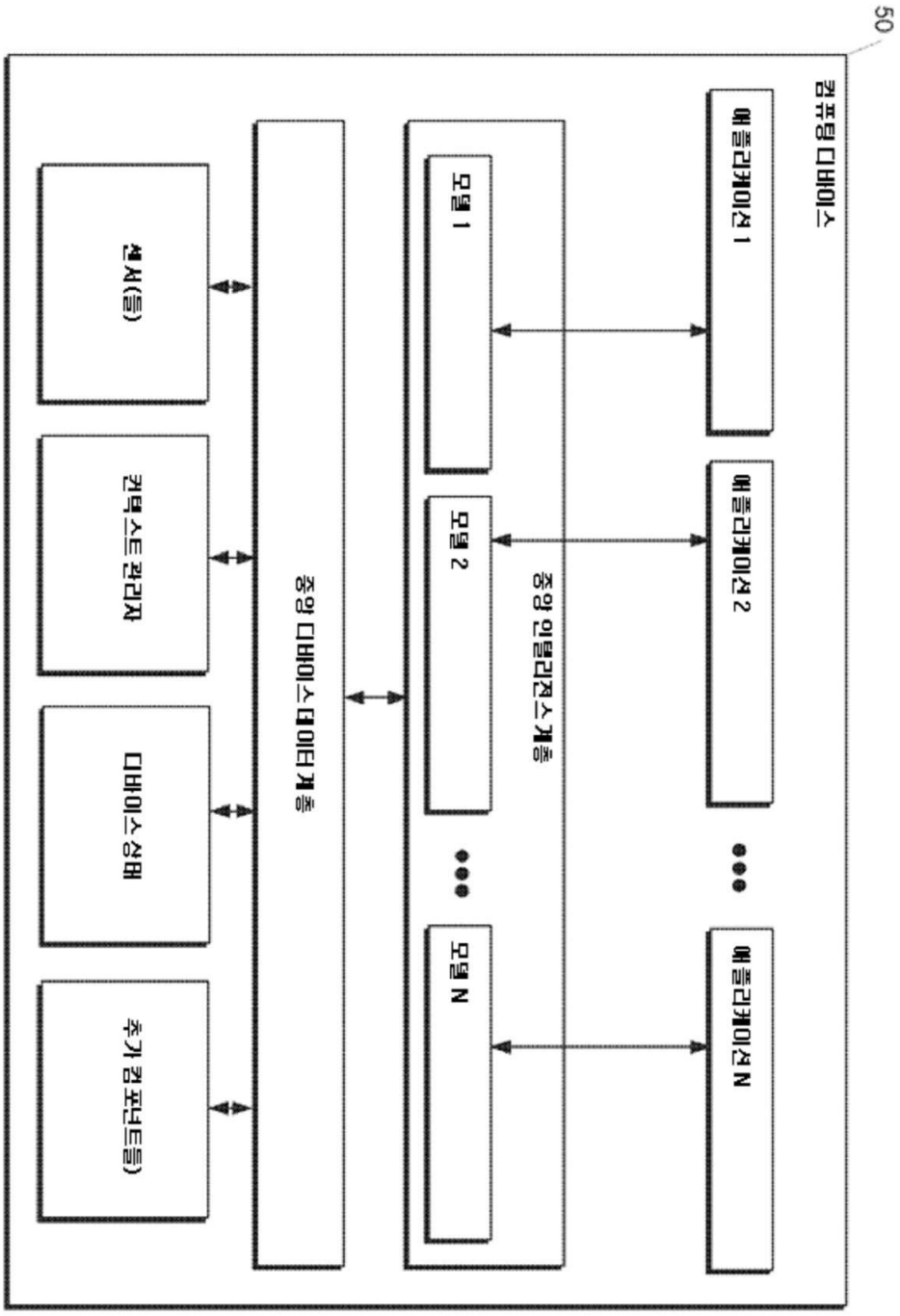
도면1a


100
102
사용자 컴퓨팅 디바이스
112
프로세서(들)
114
메모리
116
데이터
118
명령들
120
정보 추출 모델(들)
122
사용자 입력 컴포넌트
180
130
서버 컴퓨팅 시스템
132
프로세서(들)
134
메모리
136
데이터
138
명령들
140
정보 추출 모델(들)
150
트레이닝 컴퓨팅 시스템
152
프로세서(들)
154
메모리
156
데이터
158
명령들
160
모델 트레이너
162
트레이닝 데이터

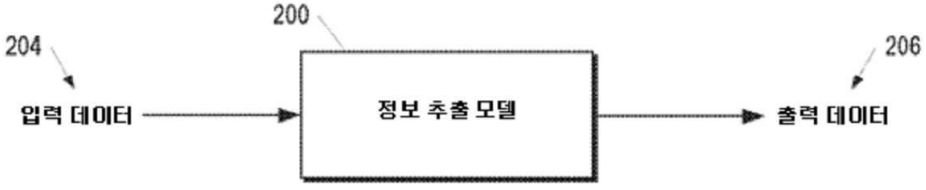
도면1b


10
컴퓨팅 디바이스
애플리케이션 1
기계 학습 라이브러리 1
기계 학습 모델 1
애플리케이션 2
기계 학습 라이브러리 2
기계 학습 모델 2
애플리케이션 N
기계 학습 라이브러리 N
기계 학습 모델 N
센서(들)
컨텍스트 관리자
디바이스 상태
추가 컴포넌트들)

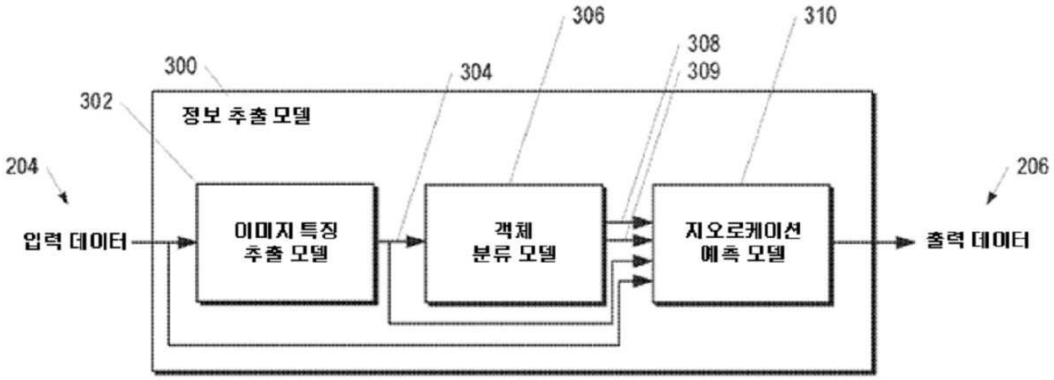
도면1c


50
컴퓨팅 디바이스
애플리케이션 1
애플리케이션 2
애플리케이션 N
모듈 1
모듈 2
모듈 N
중앙 인텔리전스 계층
중앙 디바이스 데이터 계층
센서(들)
컨텍스트 관리자
디바이스 상태
추가 컴포넌트(들)

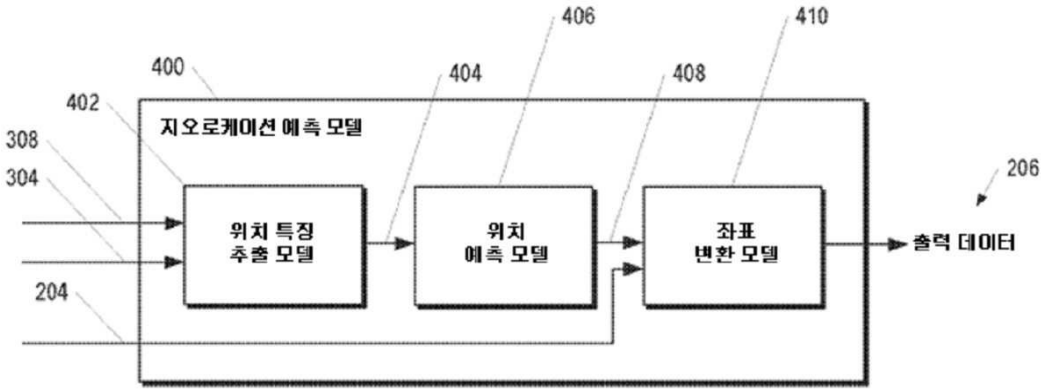
도면2


200
204
206
입력 데이터
정보 추출 모델
출력 데이터

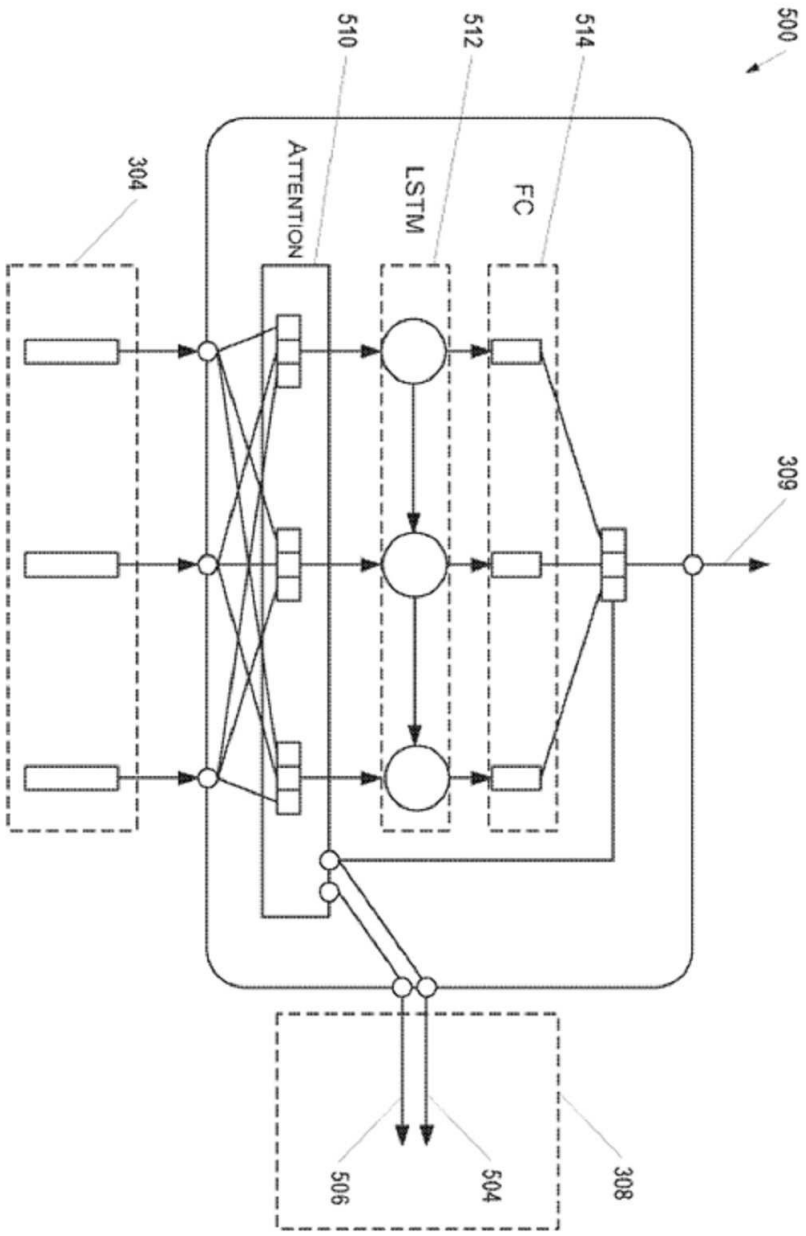
도면3


306
310
308
309
300
304
302
정보 추출 모델
204
206
입력 데이터
이미지 특징
추출 모델
객체
분류 모델
지오로케이션
예측 모델
출력 데이터

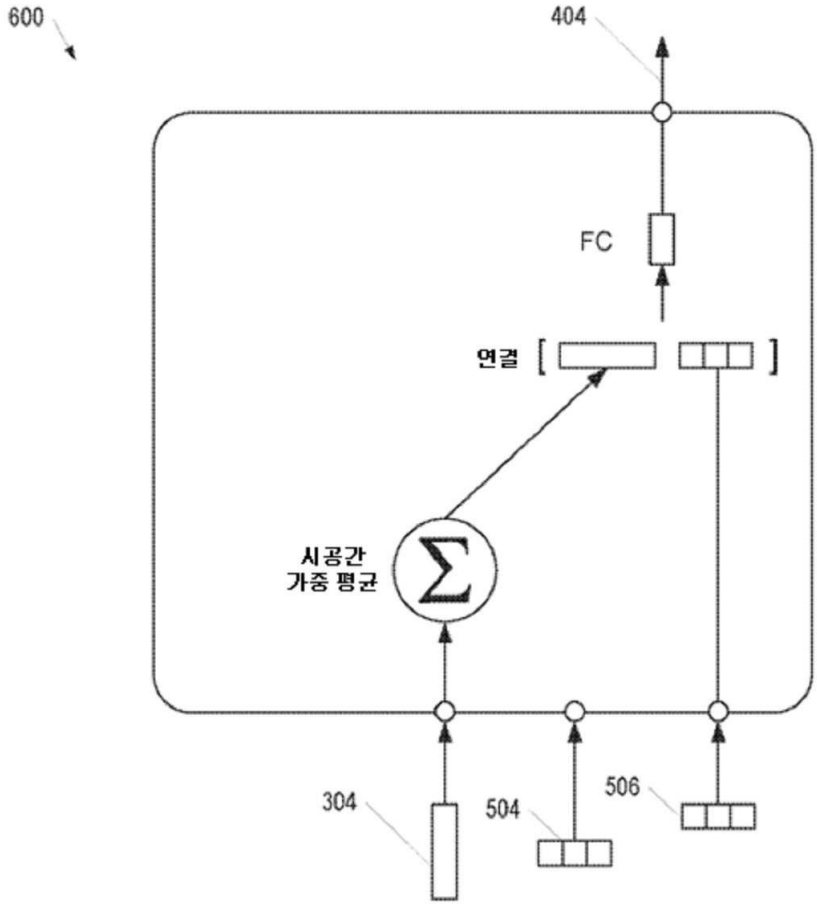
도면4


406
410
400
404
408
402
지오로케이션 예측 모델
308
304
206
위치 특징
추출 모델
위치
예측 모델
좌표
변환 모델
204
출력 데이터

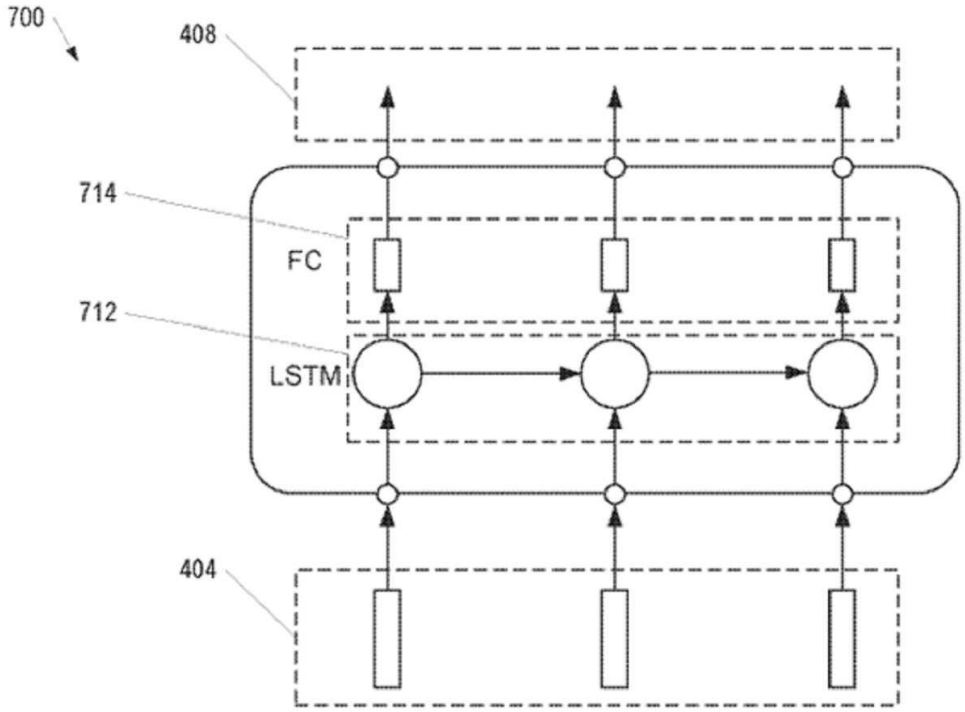
도면5



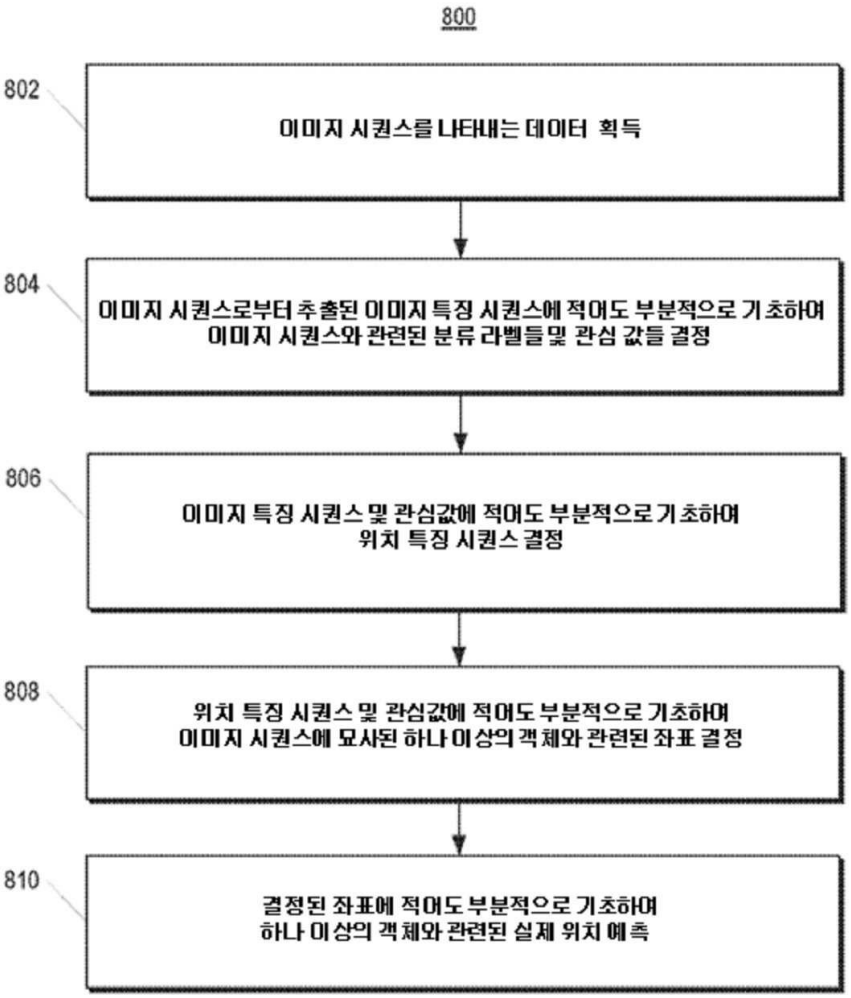
도면6



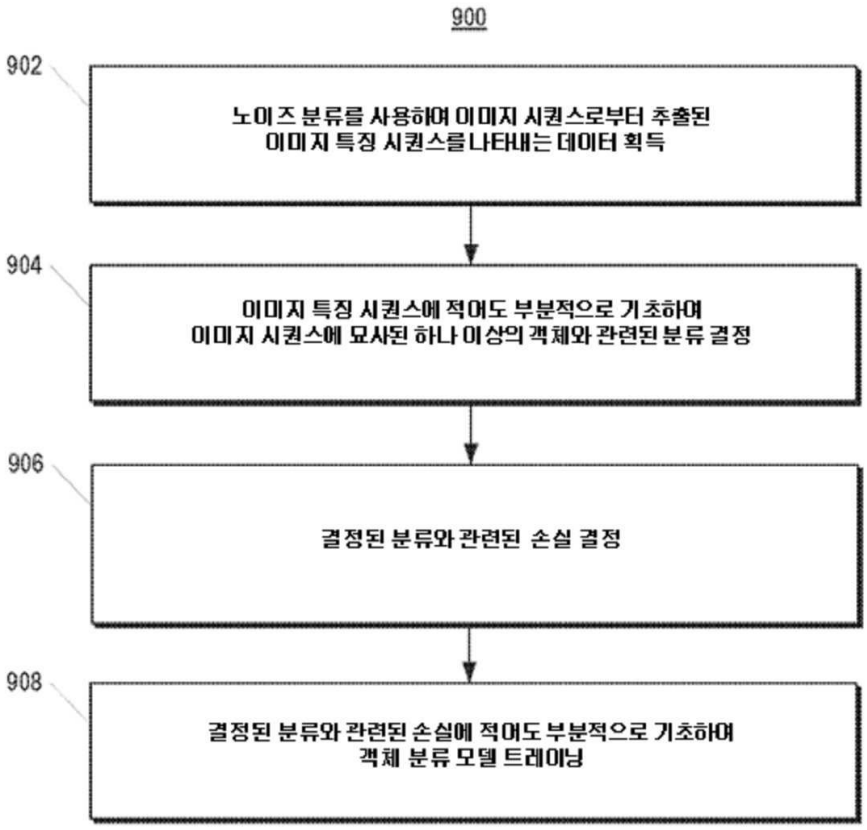
도면7



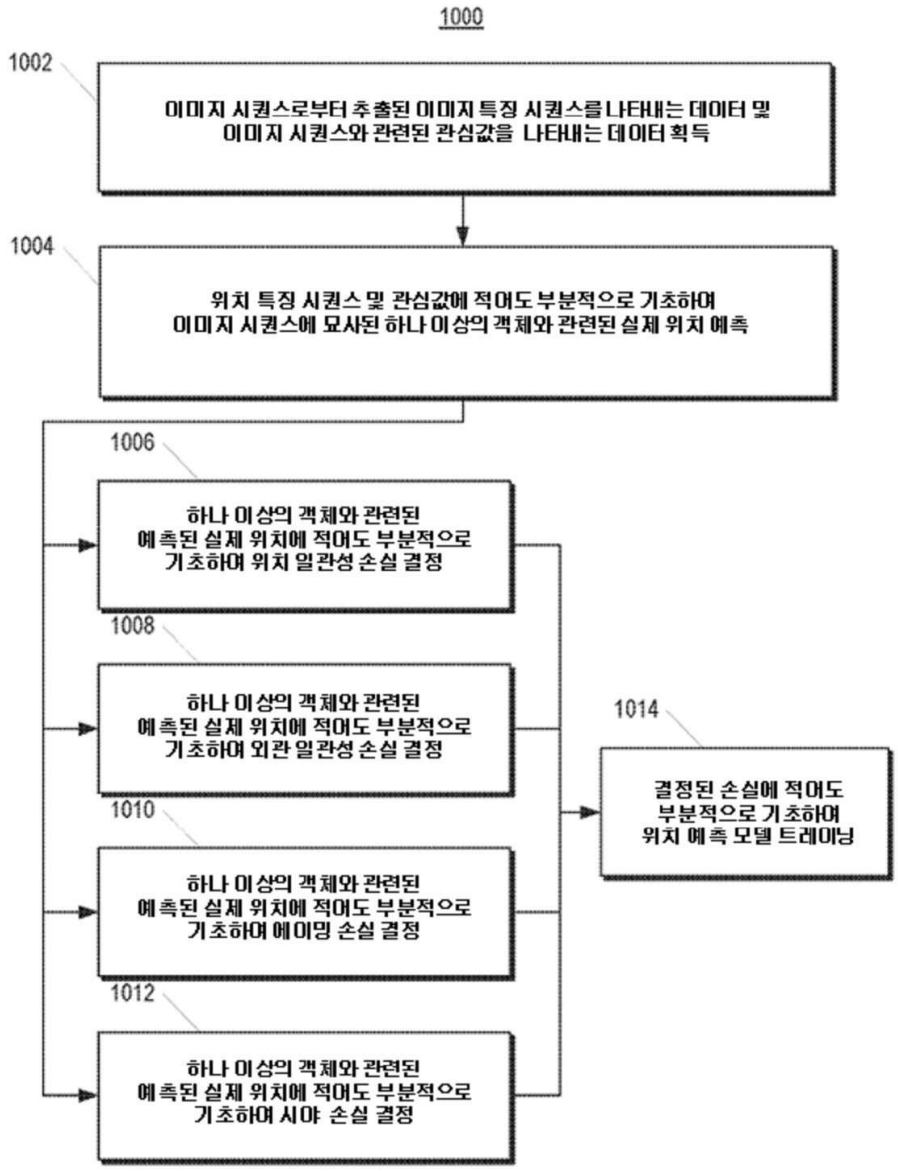
도면8


800
802
이미지 시퀀스를 나타내는 데이터 획득
804
이미지 시퀀스로부터 추출된 이미지 특징 시퀀스에 적어도 부분적으로 기초하여
이미지 시퀀스와 관련된 분류 라벨들 및 관심 값들 결정
806
이미지 특징 시퀀스 및 관심값에 적어도 부분적으로 기초하여
위치 특징 시퀀스 결정
808
위치 특징 시퀀스 및 관심값에 적어도 부분적으로 기초하여
이미지 시퀀스에 묘사된 하나 이상의 객체와 관련된 좌표 결정
810
결정된 좌표에 적어도 부분적으로 기초하여
하나 이상의 객체와 관련된 실제 위치 예측

도면9


900
902
노이즈 분류를 사용하여 이미지 시퀀스로부터 추출된
이미지 특징 시퀀스를 나타내는 데이터 획득
904
이미지 특징 시퀀스에 적어도 부분적으로 기초하여
이미지 시퀀스에 묘사된 하나 이상의 객체와 관련된 분류 결정
906
결정된 분류와 관련된 손실 결정
908
결정된 분류와 관련된 손실에 적어도 부분적으로 기초하여
객체 분류 모델 트레이닝

도면10


1000
1002
이미지 시퀀스로부터 추출된 이미지 특징 시퀀스를 나타내는 데이터 및
이미지 시퀀스와 관련된 관심값을 나타내는 데이터 획득
1004
위치 특징 시퀀스 및 관심값에 적어도 부분적으로 기초하여
이미지 시퀀스에 묘사된 하나 이상의 객체와 관련된 실제 위치 예측
1006
하나 이상의 객체와 관련된
예측된 실제 위치에 적어도 부분적으로
기초하여 위치 일관성 손실 결정
1008
하나 이상의 객체와 관련된
예측된 실제 위치에 적어도 부분적으로
기초하여 외관 일관성 손실 결정
1010
하나 이상의 객체와 관련된
예측된 실제 위치에 적어도 부분적으로
기초하여 에이밍 손실 결정
1012
하나 이상의 객체와 관련된
예측된 실제 위치에 적어도 부분적으로
기초하여 시야 손실 결정
1014
결정된 손실에 적어도
부분적으로 기초하여
위치 예측 모델 트레이닝