



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0077942
(43) 공개일자 2020년07월01일

(51) 국제특허분류(Int. Cl.)

G06T 7/246 (2017.01)

(52) CPC특허분류

G06T 7/246 (2017.01)

G06T 2207/20081 (2013.01)

(21) 출원번호 10-2018-0167361

(22) 출원일자 2018년12월21일

심사청구일자 2018년12월21일

(71) 출원인

네이버랩스 주식회사

경기도 성남시 분당구 구미로 8, 4층(구미동, 분당엠평타워)

고려대학교 산학협력단

서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)

(72) 발명자

정성균

경기도 성남시 분당구 구미로 8 분당 엠평타워 3층

김경래

서울특별시 성북구 안암로 고려대학교 자연계캠퍼스 신공학관 405호

김창수

서울특별시 성북구 안암로 고려대학교 자연계캠퍼스 공학관 508호

(74) 대리인

양성보

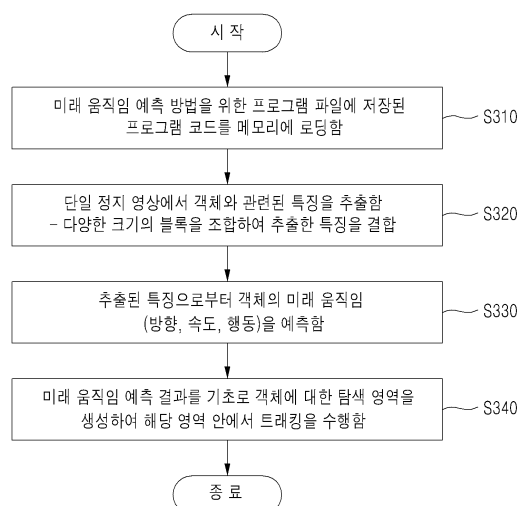
전체 청구항 수 : 총 12 항

(54) 발명의 명칭 객체의 미래 움직임 예측을 통한 동영상에서의 객체 추적을 위한 방법 및 시스템

(57) 요약

객체의 미래 움직임 예측을 통한 동영상에서의 객체 추적을 위한 방법 및 시스템이 개시된다. 객체 추적 방법은, 동영상의 일 프레임에 해당되는 정지 영상에서 상기 정지 영상에 포함된 객체와 상기 객체의 주변 정보의 특징(feature)을 추출하는 단계; 상기 추출된 특징으로부터 상기 객체의 미래 움직임(future motion)을 예측하는 단계; 및 상기 미래 움직임의 예측 결과를 기초로 다음 프레임에서의 탐색 영역을 생성하여 상기 객체의 트래킹을 수행하는 단계를 포함한다.

대표도 - 도3



명세서

청구범위

청구항 1

컴퓨터 시스템에서 수행되는 객체 추적 방법에 있어서,

동영상의 일 프레임에 해당되는 정지 영상에서 상기 정지 영상에 포함된 객체와 상기 객체의 주변 정보의 특징(feature)을 추출하는 단계;

상기 추출된 특징으로부터 상기 객체의 미래 움직임(future motion)을 예측하는 단계; 및

상기 미래 움직임의 예측 결과를 기초로 다음 프레임에서의 탐색 영역을 생성하여 상기 객체의 트래킹을 수행하는 단계

를 포함하는 객체 추적 방법.

청구항 2

제1항에 있어서,

상기 예측하는 단계는,

상기 객체의 방향(direction), 속도(velocity), 및 행동(action) 중 적어도 하나를 각 클래스로 한 인스턴스 레벨의 미래 움직임을 예측하고,

상기 수행하는 단계는,

상기 객체의 미래 움직임으로 예측된 방향, 속도, 행동 중 적어도 하나에 대응하여 상기 다음 프레임에서 이동 확률이 높은 지역을 우선적으로 탐색함으로써 상기 객체의 트래킹을 수행하는 것

을 특징으로 하는 객체 추적 방법.

청구항 3

제1항에 있어서,

상기 수행하는 단계는,

상기 객체의 미래 움직임으로 예측된 속도에 따라 상기 탐색 영역을 줄이는 단계

를 포함하는 객체 추적 방법.

청구항 4

제1항에 있어서,

상기 수행하는 단계는,

상기 객체의 미래 움직임으로 예측된 방향에 대응되는 부채꼴 모양의 탐색 영역을 결정하는 단계

를 포함하는 객체 추적 방법.

청구항 5

제1항에 있어서,

상기 수행하는 단계는,

객체 주변의 배경 이동을 보정한 후 상기 객체의 트래킹을 수행하는 것

을 특징으로 하는 객체 추적 방법.

청구항 6

제1에 있어서,

상기 예측하는 단계는,

상기 정지 영상에 포함된 객체 각각의 미래 움직임으로 예측된 방향을 클러스터링하여 각 클러스터의 그룹 방향을 예측하는 단계

를 포함하는 객체 추적 방법.

청구항 7

제1항에 있어서,

상기 추출하는 단계는,

학습 이미지에 포함된 객체와 객체의 주변 정보를 통합하여 학습한 학습 모델을 통해 상기 특징을 추출하는 것을 특징으로 하는 객체 추적 방법.

청구항 8

제1항에 있어서,

상기 추출하는 단계는,

객체 인스턴스에 대해 정답 데이터로서 미래 움직임을 나타내는 방향, 속도, 및 행동 중 적어도 하나의 속성이 할당된 이미지들을 학습한 학습 모델을 통해 상기 특징을 추출하는 것

을 특징으로 하는 객체 추적 방법.

청구항 9

제1항에 있어서,

상기 추출하는 단계는,

학습 모델을 구성하는 복수 개의 RoI(region of interest) 풀링 레이어(pooling layer)를 통해 상기 정지 영상에서 객체 영역에 대한 객체 피쳐, 객체를 포함하는 주변 영역에 대한 로컬 피쳐, 및 영상 전체 영역에 대한 글로벌 피쳐를 추출하는 것

을 특징으로 하는 객체 추적 방법.

청구항 10

제9항에 있어서,

상기 예측하는 단계는,

상기 학습 모델을 구성하는 완전 연결 레이어(fully connected layer)와 소프트맥스 레이어(softmax layer)를 통해 상기 RoI 풀링 레이어의 출력을 받아 상기 객체의 미래 움직임을 나타내는 방향, 속도, 행동 중 적어도 하나를 예측하는 것

을 특징으로 하는 객체 추적 방법.

청구항 11

제1항 내지 제10항 중 어느 한 항의 객체 추적 방법을 컴퓨터에 실행시키기 위한 프로그램이 기록되어 있는 것을 특징으로 하는 컴퓨터에서 판독 가능한 기록매체.

청구항 12

컴퓨터 시스템에 있어서,

컴퓨터에서 판독 가능한 명령을 실행하도록 구현되는 적어도 하나의 프로세서를 포함하고,

상기 적어도 하나의 프로세서는,

동영상의 일 프레임에 해당되는 정지 영상에서 상기 정지 영상에 포함된 객체와 상기 객체의 주변 정보의 특징을 추출하는 특징 추출부;

상기 추출된 특징으로부터 상기 객체의 미래 움직임을 예측하는 움직임 예측부; 및

상기 미래 움직임을 예측 결과를 기초로 다음 프레임에서의 탐색 영역을 생성하여 상기 객체의 트래킹을 수행하는 객체 트래킹부

를 포함하는 컴퓨터 시스템.

발명의 설명

기술 분야

[0001] 아래의 설명은 동영상에서 객체를 추적하는 기술에 관한 것이다.

배경 기술

[0002] 광학적 플로우(optical flow), 행동 인식(action recognition), 미래 프레임 예측(future frame prediction), 비디오 압축(video compression) 등과 같은 많은 시각 및 영상 문제에서 객체에 대한 움직임(motion)의 이해와 표현이 연구되고 있다.

[0003] 예컨대, 한국 공개특허공보 제10-2012-0106279호(공개일 2012년 09월 26일)에는 블록 매칭(block matching) 기술을 바탕으로 영상 신호에서 움직임을 예측하는 기술이 개시되어 있다.

[0004] 대부분의 기존 기술은 인접한 프레임에서 픽셀값의 차이를 최소화 하는 영역을 찾는 방식으로 움직임을 모델링하게 된다. 이는 영상을 취득하는 장비가 대부분 프레임 기반의 카메라이고 동일한 사물이 갖는 화소의 값이 유사하다는 밝기 일관성 제약(brightness consistency constraint)으로 근사된 모델이기 때문에 인간의 인지 능력에 미치지 못한다.

[0005] 반면에, 인간은 정지 영상임에도 불구하고 사물의 미래 움직임을 빠르고 정확하게 예측한다. 이는 인간이 움직임을 인식하는 과정에서 사물과 사물의 주변부의 정보를 종합적으로 판단하기 때문이라 할 수 있다.

[0006] 한편, 현존하는 대부분의 트래킹 기술은 객체 검출 결과를 중심으로 주변 영역에 대한 완전 탐색(exhaustive search)을 통해 매칭 스코어가 가장 높은 위치로 객체가 이동했다고 판단한다.

[0007] 기존 트래킹 기술은 인접한 프레임을 분석하여 객체의 움직임을 예측하기 때문에, 입력 영상의 프레임 레이트(frame rate)에 따라 트래킹 성능(매칭 정확도, 속도 측면 등)이 가변적이며 통신 지연 등으로 인해 갑작스럽게 입력 영상의 프레임 저하가 발생하는 경우 안정적인 트래킹 결과를 보장할 수 없다.

발명의 내용

해결하려는 과제

[0008] 단일 정지 영상(single still image)에서 객체의 미래 움직임을 예측하여 객체를 안정적으로 추적할 수 있는 방법과 시스템을 제공한다.

[0009] 프레임 레이트가 현저히 낮은 동영상이나 프레임 드롭(frame drop)이 발생하는 경우 트래킹 성능 저하를 예방하고 안정적인 트래킹을 보장할 수 있는 방법과 시스템을 제공한다.

[0010] 단일 정지 영상인 일 프레임에서 객체의 미래 움직임을 예측하여 예측 결과를 기반으로 다음 프레임에서 객체를 탐색함으로써 객체 탐색 영역을 효율적으로 줄일 수 있는 방법과 시스템을 제공한다.

과제의 해결 수단

- [0011] 컴퓨터 시스템에서 수행되는 객체 추적 방법에 있어서, 동영상의 일 프레임에 해당되는 정지 영상에서 상기 정지 영상에 포함된 객체와 상기 객체의 주변 정보의 특징(feature)을 추출하는 단계; 상기 추출된 특징으로부터 상기 객체의 미래 움직임(future motion)을 예측하는 단계; 및 상기 미래 움직임의 예측 결과를 기초로 다음 프레임에서의 탐색 영역을 생성하여 상기 객체의 트래킹을 수행하는 단계를 포함하는 객체 추적 방법을 제공한다.
- [0012] 일 측면에 따르면, 상기 예측하는 단계는, 상기 객체의 방향(direction), 속도(velocity), 및 행동(action) 중 적어도 하나를 각 클래스로 한 인스턴스 레벨의 미래 움직임을 예측하고, 상기 수행하는 단계는, 상기 객체의 미래 움직임으로 예측된 방향, 속도, 행동 중 적어도 하나에 대응하여 상기 다음 프레임에서 이동 확률이 높은 지역을 우선적으로 탐색함으로써 상기 객체의 트래킹을 수행할 수 있다.
- [0013] 다른 측면에 따르면, 상기 수행하는 단계는, 상기 객체의 미래 움직임으로 예측된 속도에 따라 상기 탐색 영역을 줄이는 단계를 포함할 수 있다.
- [0014] 또 다른 측면에 따르면, 상기 수행하는 단계는, 상기 객체의 미래 움직임으로 예측된 방향에 대응되는 부채꼴 모양의 탐색 영역을 결정하는 단계를 포함할 수 있다.
- [0015] 또 다른 측면에 따르면, 상기 수행하는 단계는, 객체 주변의 배경 이동을 보정한 후 상기 객체의 트래킹을 수행할 수 있다.
- [0016] 또 다른 측면에 따르면, 상기 예측하는 단계는, 상기 정지 영상에 포함된 객체 각각의 미래 움직임으로 예측된 방향을 클러스터링하여 각 클러스터의 그룹 방향을 예측하는 단계를 포함할 수 있다.
- [0017] 또 다른 측면에 따르면, 상기 추출하는 단계는, 학습 이미지에 포함된 객체와 객체의 주변 정보를 통합하여 학습한 학습 모델을 통해 상기 특징을 추출할 수 있다.
- [0018] 또 다른 측면에 따르면, 상기 추출하는 단계는, 객체 인스턴스에 대해 정답 데이터로서 미래 움직임을 나타내는 방향, 속도, 및 행동 중 적어도 하나의 속성이 할당된 이미지들을 학습한 학습 모델을 통해 상기 특징을 추출할 수 있다.
- [0019] 또 다른 측면에 따르면, 상기 추출하는 단계는, 학습 모델을 구성하는 복수 개의 RoI(region of interest) 풀링 레이어(pooling layer)를 통해 상기 정지 영상에서 객체 영역에 대한 객체 피쳐, 객체를 포함하는 주변 영역에 대한 로컬 피쳐, 및 영상 전체 영역에 대한 글로벌 피쳐를 추출할 수 있다.
- [0020] 또 다른 측면에 따르면, 상기 예측하는 단계는, 상기 학습 모델을 구성하는 완전 연결 레이어(fully connected layer)와 소프트맥스 레이어(softmax layer)를 통해 상기 RoI 풀링 레이어의 출력을 받아 상기 객체의 미래 움직임을 나타내는 방향, 속도, 행동 중 적어도 하나를 예측할 수 있다.
- [0021] 상기 객체 추적 방법을 컴퓨터에 실행시키기 위한 프로그램이 기록되어 있는 것을 특징으로 하는 컴퓨터에서 판독 가능한 기록매체를 제공한다.
- [0022] 컴퓨터 시스템에 있어서, 컴퓨터에서 판독 가능한 명령을 실행하도록 구현되는 적어도 하나의 프로세서를 포함하고, 상기 적어도 하나의 프로세서는, 동영상의 일 프레임에 해당되는 정지 영상에서 상기 정지 영상에 포함된 객체와 상기 객체의 주변 정보의 특징을 추출하는 특징 추출부; 상기 추출된 특징으로부터 상기 객체의 미래 움직임을 예측하는 움직임 예측부; 및 상기 미래 움직임의 예측 결과를 기초로 다음 프레임에서의 탐색 영역을 생성하여 상기 객체의 트래킹을 수행하는 객체 트래킹부를 포함하는 컴퓨터 시스템을 제공한다.

발명의 효과

- [0023] 본 발명의 실시예들에 따르면, 단일 정지 영상에서 객체의 미래 움직임을 예측하여 객체를 안정적으로 추적할 수 있다.
- [0024] 본 발명의 실시예들에 따르면, 단일 정지 영상에서 객체의 미래 움직임을 예측할 수 있어 프레임 레이트가 현저히 낮은 동영상이나 프레임 드롭이 발생하는 경우에도 트래킹 성능 저하를 예방하고 안정적인 트래킹을 보장할 수 있다.
- [0025] 본 발명의 실시예들에 따르면, 단일 정지 영상인 일 프레임에서 객체의 미래 움직임을 예측하여 예측 결과를 기반으로 다음 프레임에서 객체를 탐색함으로써 객체 탐색 영역을 효율적으로 줄일 수 있다.

도면의 간단한 설명

- [0026] 도 1은 본 발명의 일실시예에 있어서 컴퓨터 시스템의 내부 구성의 일례를 설명하기 위한 블록도이다.
- 도 2는 본 발명의 일실시예에 따른 컴퓨터 시스템의 프로세서가 포함할 수 있는 구성요소의 예를 도시한 도면이다.
- 도 3은 본 발명의 일실시예에 따른 컴퓨터 시스템이 수행할 수 있는 객체 추적 방법의 예를 도시한 순서도이다.
- 도 4는 본 발명의 일실시예에 있어서 학습 이미지로 사용 가능한 이미지 예시를 설명하기 위한 도면이다.
- 도 5는 본 발명의 일실시예에 있어서 움직임에 대한 속도 클래스와 행동 클래스의 예시를 도시한 것이다.
- 도 6은 본 발명의 일실시예에 있어서 미래 움직임 예측을 위한 DNN 구조를 도시한 것이다.
- 도 7은 MCRoI 풀링 레이어를 포함하는 DNN의 분류 성능을 설명하기 위한 테이블을 도시한 것이다.
- 도 8은 본 발명의 일실시예에 있어서 미래 움직임 예측 결과를 이용하여 객체 트래킹을 위한 탐색 영역을 생성하는 과정을 설명하기 위한 예시 도면이다.
- 도 9는 본 발명의 일실시예에 있어서 미래 움직임 예측 결과를 이용한 군중 분석 결과의 예시를 도시한 것이다.

발명을 실시하기 위한 구체적인 내용

- [0027] 이하, 본 발명의 실시예를 첨부된 도면을 참조하여 상세하게 설명한다.
- [0029] 본 발명의 실시예들은 단일 정지 영상에서 객체의 미래 움직임을 예측하는 방식을 통해 동영상에서의 객체를 추적하는 기술에 관한 것이다.
- [0030] 본 명세서에서 구체적으로 개시되는 것들을 포함하는 실시예들은 프레임 레이트가 현저히 낮은 동영상이나 프레임 드롭이 발생하더라도 객체를 안정적으로 추적할 수 있고, 이를 통해 효율성, 정확성, 신속성, 비용 절감 등의 측면에 있어서 상당한 장점들을 달성한다.
- [0031] 도 1은 본 발명의 일실시예에 있어서 컴퓨터 시스템의 내부 구성의 일례를 설명하기 위한 블록도이다. 예를 들어, 본 발명의 실시예들에 따른 객체 추적 시스템이 도 1의 컴퓨터 시스템(100)을 통해 구현될 수 있다. 도 1에 도시한 바와 같이, 컴퓨터 시스템(100)은 객체 추적 방법을 실행하기 위한 구성요소로서 프로세서(110), 메모리(120), 영구 저장 장치(130), 버스(140), 입출력 인터페이스(150) 및 네트워크 인터페이스(160)를 포함할 수 있다.
- [0032] 프로세서(110)는 동영상에서 객체를 추적하기 위한 구성요소로서 명령어들의 시퀀스를 처리할 수 있는 임의의 장치를 포함하거나 그의 일부일 수 있다. 프로세서(110)는 예를 들어 컴퓨터 프로세서, 이동 장치 또는 다른 전자 장치 내의 프로세서 및/또는 디지털 프로세서를 포함할 수 있다. 프로세서(110)는 예를 들어, 서버 컴퓨팅 디바이스, 서버 컴퓨터, 일련의 서버 컴퓨터들, 서버 팜, 클라우드 컴퓨터, 콘텐츠 플랫폼 등에 포함될 수 있다. 프로세서(110)는 버스(140)를 통해 메모리(120)에 접속될 수 있다.
- [0033] 메모리(120)는 컴퓨터 시스템(100)에 의해 사용되거나 그에 의해 출력되는 정보를 저장하기 위한 휘발성 메모리, 영구, 가상 또는 기타 메모리를 포함할 수 있다. 메모리(120)는 예를 들어 랜덤 액세스 메모리(RAM: random access memory) 및/또는 다이내믹 RAM(DRAM: dynamic RAM)을 포함할 수 있다. 메모리(120)는 컴퓨터 시스템(100)의 상태 정보와 같은 임의의 정보를 저장하는 데 사용될 수 있다. 메모리(120)는 예를 들어 객체 추적을 위한 명령어들을 포함하는 컴퓨터 시스템(100)의 명령어들을 저장하는 데에도 사용될 수 있다. 컴퓨터 시스템(100)은 필요에 따라 또는 적절한 경우에 하나 이상의 프로세서(110)를 포함할 수 있다.
- [0034] 버스(140)는 컴퓨터 시스템(100)의 다양한 컴포넌트들 사이의 상호작용을 가능하게 하는 통신 기반 구조를 포함할 수 있다. 버스(140)는 예를 들어 컴퓨터 시스템(100)의 컴포넌트들 사이에, 예를 들어 프로세서(110)와 메모리(120) 사이에 데이터를 운반할 수 있다. 버스(140)는 컴퓨터 시스템(100)의 컴포넌트들 간의 무선 및/또는 유선 통신 매체를 포함할 수 있으며, 병렬, 직렬 또는 다른 토폴로지 배열들을 포함할 수 있다.
- [0035] 영구 저장 장치(130)는 (예를 들어, 메모리(120)에 비해) 소정의 연장된 기간 동안 데이터를 저장하기 위해 컴퓨터 시스템(100)에 의해 사용되는 바와 같은 메모리 또는 다른 영구 저장 장치와 같은 컴포넌트들을 포함할 수

있다. 영구 저장 장치(130)는 컴퓨터 시스템(100) 내의 프로세서(110)에 의해 사용되는 바와 같은 비휘발성 메인 메모리를 포함할 수 있다. 영구 저장 장치(130)는 예를 들어 플래시 메모리, 하드 디스크, 광 디스크 또는 다른 컴퓨터 판독 가능 매체를 포함할 수 있다.

[0036] 입출력 인터페이스(150)는 키보드, 마우스, 음성 명령 입력, 디스플레이 또는 다른 입력 또는 출력 장치에 대한 인터페이스들을 포함할 수 있다. 구성 명령들 및/또는 객체 추적을 위한 입력이 입출력 인터페이스(150)를 통해 수신될 수 있다.

[0037] 네트워크 인터페이스(160)는 근거리 네트워크 또는 인터넷과 같은 네트워크들에 대한 하나 이상의 인터페이스를 포함할 수 있다. 네트워크 인터페이스(160)는 유선 또는 무선 접속들에 대한 인터페이스들을 포함할 수 있다. 구성 명령들 및/또는 객체 추적을 위한 입력이 네트워크 인터페이스(160)를 통해 수신될 수 있다.

[0038] 또한, 다른 실시예들에서 컴퓨터 시스템(100)은 도 1의 구성요소들보다 더 많은 구성요소들을 포함할 수도 있다. 그러나, 대부분의 종래기술적 구성요소들을 명확하게 도시할 필요성은 없다. 예를 들어, 컴퓨터 시스템(100)은 상술한 입출력 인터페이스(150)와 연결되는 입출력 장치들 중 적어도 일부를 포함하도록 구현되거나 또는 트랜시버(transceiver), GPS(Global Positioning System) 모듈, 카메라, 각종 센서, 데이터베이스 등과 같은 다른 구성요소들을 더 포함할 수도 있다.

[0039] 본 발명에서는 인간 수준의 움직임 예측 능력을 모사하기 위해 다양한 수준의 의미 정보를 통합하는 학습 모델(일례로, DNN(deep neural network))를 제안하여 이를 통해 단일 이미지에서 인스턴스 레벨의 미래 움직임을 예측할 수 있다.

[0040] 본 명세서에서는 보행자를 대표적인 객체로 하여 해당 객체의 미래 움직임을 예측하는 실시예를 설명하고 있으나, 이에 한정되는 것은 자동차, 동물 등 다른 종류의 객체를 대상으로 적용할 수 있다.

[0041] 도 2는 본 발명의 일실시예에 따른 컴퓨터 시스템의 프로세서가 포함할 수 있는 구성요소의 예를 도시한 도면이고, 도 3은 본 발명의 일실시예에 따른 컴퓨터 시스템이 수행할 수 있는 객체 추적 예측 방법의 예를 도시한 순서도이다.

[0042] 도 2에 도시된 바와 같이, 프로세서(110)는 특징 추출부(210), 움직임 예측부(220), 및 객체 트래킹부(230)를 포함할 수 있다. 이러한 프로세서(110)의 구성요소들은 적어도 하나의 프로그램 코드에 의해 제공되는 제어 명령에 따라 프로세서(110)에 의해 수행되는 서로 다른 기능들(different functions)의 표현들일 수 있다. 예를 들어, 프로세서(110)가 정지 영상에서 객체의 특징을 추출하도록 컴퓨터 시스템(100)을 제어하기 위해 동작하는 기능적 표현으로서 특징 추출부(210)가 사용될 수 있다. 프로세서(110) 및 프로세서(110)의 구성요소들은 도 3의 객체 추적 방법이 포함하는 단계들(S310 내지 S340)을 수행할 수 있다. 예를 들어, 프로세서(110) 및 프로세서(110)의 구성요소들은 메모리(120)가 포함하는 운영체제의 코드와 상술한 적어도 하나의 프로그램 코드에 따른 명령(instruction)을 실행하도록 구현될 수 있다. 여기서, 적어도 하나의 프로그램 코드는 객체 추적 방법을 처리하기 위해 구현된 프로그램의 코드에 대응될 수 있다.

[0043] 객체 추적 방법은 도시된 순서대로 발생하지 않을 수 있으며, 단계들 중 일부가 생략되거나 추가의 과정이 더 포함될 수 있다.

[0044] 단계(S310)에서 프로세서(110)는 객체 추적 방법을 위한 프로그램 파일에 저장된 프로그램 코드를 메모리(120)에 로딩할 수 있다. 예를 들어, 객체 추적 방법을 위한 프로그램 파일은 도 1을 통해 설명한 영구 저장 장치(130)에 저장되어 있을 수 있고, 프로세서(110)는 버스를 통해 영구 저장 장치(130)에 저장된 프로그램 파일로부터 프로그램 코드가 메모리(120)에 로딩되도록 컴퓨터 시스템(110)을 제어할 수 있다. 이때, 프로세서(110) 및 프로세서(110)가 포함하는 특징 추출부(210), 움직임 예측부(220), 및 객체 트래킹부(230) 각각은 메모리(120)에 로딩된 프로그램 코드 중 대응하는 부분의 명령을 실행하여 이후 단계들(S320 내지 S340)을 실행하기 위한 프로세서(110)의 서로 다른 기능적 표현들일 수 있다. 단계들(S320 내지 S340)의 실행을 위해, 프로세서(110) 및 프로세서(110)의 구성요소들은 직접 제어 명령에 따른 연산을 처리하거나 또는 컴퓨터 시스템(100)을 제어할 수 있다.

[0045] 본 발명에 따른 객체 추적 방법은 객체 검출(object detection)을 통해 입력 동영상의 일 프레임인 단일 정지 영상에서 객체 영역을 추출하는 단계(S320); 객체 영역에 대한 미래 움직임 예측을 통해 해당 객체의 다음 방향, 속도, 행동을 미리 정해진 클래스로 예측하는 단계(S330); 및 일 프레임의 미래 움직임 예측 결과를 바탕으로 다음 프레임에서의 탐색 영역을 적응적으로 생성하여 해당 영역 안에서 객체의 트래킹을 수행하는 단계

(S340)를 포함할 수 있다.

- [0046] 단계(S320)에서 특징 추출부(210)는 동영상의 일 프레임에 해당되는 단일 정지 영상을 입력 이미지로 받아 입력 이미지에서 객체와 관련된 특징을 추출할 수 있다. 특히, 특징 추출부(210)는 입력 이미지에서 다양한 크기의 블록, 즉 객체 영역, 객체를 포함하는 주변 영역(local context), 및 영상 전체 영역(global context)의 특징을 추출한 후 추출된 세 가지 종류의 특징을 모두 결합할 수 있다.
- [0047] 단계(S330)에서 움직임 예측부(220)는 단계(S320)에서 추출된 특징으로부터 해당 객체의 미래 움직임을 예측할 수 있다. 움직임 예측부(220)는 객체의 미래 움직임을 세 가지의 움직임, 즉 방향(direction), 속도(velocity), 및 행동(action)으로 구분하여 예측할 수 있다. 움직임 예측부(220)는 단일 이미지에서 방향, 속도, 행동을 클래스로 한 인스턴스 레벨의 미래 움직임을 예측할 수 있다. 예를 들어, 방향은 8개의 방위(N, E, S, W, NE, SE, SW, NW)로 구분할 수 있으며, 이는 영상 내의 좌표를 기반으로 할 수 있다. 또한, 속도는 복수개의 단계, 예를 들어 정지(stop), 느림(slow), 빠름(fast) 3단계로 구분할 수 있다. 행동은 상황에 따라 가변적이거나, 자율주행 시스템에서 보행자를 객체로 인식하는 경우, 예를 들어 보행자의 행동을 일반 보행(sidewalk), 횡단보도 보행(crosswalk), 무단횡단 보행(jaywalk)으로 구분할 수 있다.
- [0048] 단계(S340)에서 객체 트래킹부(230)는 단계(S330)에서 예측된 객체의 미래 움직임을 기초로 해당 객체에 대한 탐색 영역을 생성하여 해당 영역 안에서 트래킹을 수행할 수 있다. 객체 트래킹부(230)는 이전 프레임인 단일 정지 영상에서 객체의 미래 움직임으로 예측된 결과를 기반으로 해당 객체가 다음 프레임에서 이동할 확률이 높은 지역을 우선적으로 탐색할 수 있다. 객체 트래킹부(230)는 객체의 미래 움직임 예측 결과에 따라 객체에 대한 탐색 지점(search point)의 형태를 다르게 하여 속도 성능 저하 없이 객체 트래킹을 안정적으로 수행할 수 있다.
- [0049] 단일 정지 영상에서 객체의 미래 움직임을 예측할 수 있어 프레임 레이트가 1fps 이하로 현저히 낮은 동영상에서도 객체를 안정적으로 트래킹할 수 있고, 자율 주행차나 로봇 등이 주변 환경을 인식하는 과정에서 갑작스러운 프레임 드롭이 발생하더라도 객체 트래킹 성능이 저하되는 것을 예방할 수 있고 주변 객체들의 행동을 미리 예측할 수 있다.
- [0050] 본 발명에서는 단일 이미지에서 객체의 방향, 속도, 행동을 각 클래스로 한 인스턴스의 미래 움직임을 예측할 수 있으며, 이러한 미래 움직임의 예측을 수행하기 위해 다양한 수준의 의미 정보를 통합하는 DNN 모델을 제안한다.
- [0051] 도 4를 참조하면, 인간은 정지된 하나의 이미지(400)를 보면 이미지(400)에 포함된 객체(401)의 다음 움직임을 어느 정도 예측이 가능하다. 본 발명은 미래의 움직임에 관한 인간 수준의 움직임 예측 능력을 구현하기 위해 DNN을 이용하고자 한다.
- [0052] 이하에서는 많은 비전 어플리케이션에서 가장 관심 있는 대상인 보행자 인스턴스를 대표적인 예시로 하여 구체적인 실시예를 설명하기로 한다.
- [0053] 기존 데이터셋이나 인터넷 상의 이미지 데이터베이스 시스템 상에서 보행자가 포함된 이미지를 학습 이미지로 선택하여 미래 움직임 데이터셋을 수집할 수 있다. 이때, 학습 이미지에 포함된 보행자 영역에 대해 경계 박스 어노테이션(bounding box annotation)을 적용한 다음, 각 보행자 인스턴스에 정답 데이터로서 미래 움직임의 세 가지 속성(즉, 방향, 속도, 행동)을 수동으로 할당한다.
- [0054] 방향 클래스는 8개의 방향(N, E, S, W, NE, SE, SW, NW)으로 정의될 수 있다. 그리고, 도 5를 참조하면 속도 클래스(510)는 정지, 느림, 빠름으로 정의할 수 있고, 행동 클래스(520)는 일반 보행, 횡단보도 보행, 무단횡단 보행으로 정의할 수 있다. 이는 하나의 예시일 뿐, 경우에 따라 방향, 속도, 행동 클래스가 다르게 정의될 수도 있다. 학습 이미지 별로 보행자 인스턴스의 방향, 속도, 행동의 각 클래스에 수동으로 라벨을 붙이게 된다.
- [0055] 일반적으로 인간은 움직이는 물체를 볼 때 대상 물체와 그 주변 환경의 시각적 정보를 동시에 지각한다. 이와 비슷하게, 본 발명에서는 미래 움직임 예측을 위한 장면 컨텍스트(scene context)를 활용하기 위해, 객체, 로컬 및 글로벌 피처를 통합하는 MCRoI(multi-context region of interest) 풀링 레이어(pooling layer)를 제안한다. MCRoI 풀링 레이어를 DenseNet(densely connected convolutional network)에 통합하여 인스턴스의 미래 방향, 속도 및 행동을 예측하기 위한 통합 모델을 학습한다.
- [0056] 본 발명에서는 방향, 속도, 행동에 대한 세 가지 분류 작업을 수행하여 보행자의 미래 움직임을 예측하기 위해 DNN을 구축한다.

- [0057] 도 6은 미래 움직임 예측을 위한 DNN 구조를 도시한 것이다.
- [0058] 도 6을 참조하면, DNN(600)은 입력 이미지 패치를 처리하게 되는데, 이때 입력 이미지 패치는 보행자가 중앙에 위치하며 세 가지 분류 결과를 생성한다.
- [0059] DNN(600)은 컴퓨터 시스템(100)의 프로세서(110)에 포함되는 것으로, 특징 추출부(210)와, 움직임 예측부(220)로 구성될 수 있다.
- [0060] 보행자 영역(61)에 대해 높이가 h인 경계 박스가 어노테이션된 학습 이미지에서 보행자 영역(61)의 경계 박스를 기준으로 $2h \times 2h$ 크기로 자른 패치를 입력 이미지(60)로 사용하는 것으로 가정한다.
- [0061] 특징 추출부(210)는 컨볼루션 레이어 및 MCRoI 풀링 레이어를 포함할 수 있다. 이때, 컨볼루션 레이어는 DenseNet으로 구성될 수 있다.
- [0062] 특징 추출부(210)는 DenseNet을 기반으로 객체 영역에 해당되는 보행자 영역(61), 객체를 포함하는 주변 영역(62), 및 영상 전체 영역(63)의 특징을 만들어 낸다. 즉, 특징 추출부(210)는 DenseNet의 출력으로 객체 피쳐, 로컬 컨텍스트 피쳐, 글로벌 컨텍스트 피쳐를 추출한다.
- [0063] 특징 추출부(210)는 4개의 RoI 풀링 레이어를 사용하는 MCRoI 풀링 레이어를 포함한다.
- [0064] 보행자 영역(61)에 대해 어노테이션된 경계 박스는 공간 해상도 7×7 에 풀링되는 객체 피쳐의 RoI에 해당된다. 이때, 객체 영역에 해당되는 보행자 영역(61)은 최소의 배경과 함께 보행자의 외관 정보를 포함한다.
- [0065] 주변 영역(62)은 경계 박스를 기준으로 $1.3h \times 1.3h$ 크기로 자른 패치로, 두 가지의 로컬 컨텍스트 피쳐를 위한 RoI에 해당된다. 패치는 각각 두 개의 RoI 풀링 레이어에 의해 해상도 9×9 및 11×11 에 풀링된다. 로컬 컨텍스트 피쳐에는 보행자 주변의 배경 정보와 보행자 외관이 포함된다.
- [0066] 영상 전체 영역(63)에 해당되는 $2h \times 2h$ 크기의 패치는 글로벌 컨텍스트 피쳐를 위한 RoI에 해당되며, 해상도 5×5 에 풀링된다. 글로벌 컨텍스트 피쳐는 장면에 대한 전체적인 의미 정보를 제공하기 때문에 미래 움직임의 예측에 있어 중요하다.
- [0067] 4개의 RoI 풀링 레이어의 출력 사이즈는 경험적으로 결정된다.
- [0068] 움직임 예측부(220)는 완전 연결 레이어(fully connected layer)와 소프트맥스 레이어(softmax layer)로 구성될 수 있다. 특히, 움직임 예측부(220)는 MCRoI 풀링 레이어의 출력을 받아 완전 연결 레이어의 3단계로 처리한 후 세 개의 소프트맥스 레이어로 처리하여 보행자의 방향, 속도, 및 행동을 각각 분류한다.
- [0069] DNN의 학습 과정은 다음과 같다.
- [0070] MCRoI 풀링 레이어는 4개의 풀링 레이어를 사용하여 컨볼루션 레이어와 완전 연결 레이어를 연결한다.
- [0071] 도 6의 DNN(600)은 임의의 사이즈의 입력 패치를 적용할 수 있으나, 효과적인 학습 및 추론을 위해 입력 패치의 크기를 400×400 으로 표준화하여 크기가 200픽셀인 보행자 영역(61)을 포함한다.
- [0072] 전체 손실 함수는 수학식 1과 같이 세 가지 손실의 합으로 정의된다.
- [0073] [수학식 1]
- $$L = L_{Dir} + L_{Vel} + L_{Act}$$
- [0074]
- [0075] 여기서, L_{Dir} 은 방향 분류에 대한 손실을, L_{Vel} 은 속도 분류에 대한 손실을, L_{Act} 는 행동 분류에 대한 손실을 의미한다.
- [0076] 각 분류 손실에 대해 크로스 엔트로피(cross entropy)를 적용한다.
- [0077] 방향 분류 손실의 경우 수학식 2와 같이 정의된다.

[0078] [수학식 2]

$$L_{\text{Dir}}(\mathbf{p}, \mathbf{q}) = - \sum_{i=0}^7 q_i \log p_i$$

[0079]

[0080] 여기서, $\mathbf{p} = \{p_i : i = 0, \dots, 7\}$ 는 8방향에 대한 소프트맥스 확률 벡터이고, $\mathbf{q} = \{q_i : i = 0, \dots, 7\}$ 는 실측 결과(ground-truth) 바이너리 벡터를 의미한다.

[0081]

L_{Vel} 과 L_{Act} 또한 수학식 2와 유사하게 정의된다.

[0082]

본 실시예에서는 0.9 모멘텀(momentum)과 14개의 에포크(epoch)에 대하여 배치 사이즈(batch size)가 4인 확률적인 경사 하강법(stochastic gradient descent)을 통해 도 6의 DNN(600)을 학습시킨다. 초기 파라미터로서 ImageNet에서 미리 학습한 DenseNet 모델을 사용한다.

[0083]

따라서, 본 발명에서는 MCRoI 풀링 레이어를 통해서 객체와 객체의 주변 정보를 통합하여 학습하는 네트워크 모델을 구현할 수 있고, 이를 통해 정지된 단일 영상만으로 객체에 대한 인스턴스 레벨의 미래 움직임(방향과 속도 및 행동)을 예측할 수 있다.

[0084]

도 7은 MCRoI 풀링 레이어를 포함하는 DNN의 분류 성능을 설명하기 위한 테이블을 도시한 것이다.

[0085]

도 7의 테이블은 MCRoI 풀링 레이어를 포함하는 네트워크 구조인 'Proposed' 네트워크의 성능을 기존 네트워크(객체 피처를 분류에 사용하는 'Object' 네트워크, 로컬 컨텍스트 피처를 분류에 사용하는 'Local context' 네트워크, 글로벌 컨텍스트 피처를 분류에 사용하는 'Global context' 네트워크, 객체 피처와 로컬 컨텍스트 피처를 분류에 사용하는 'Object+Local' 네트워크)와 비교한 것이다.

[0086]

도 7에서 'Direction+'는 인간조차 하나의 이미지를 보고 서로 다른 클래스로 판단하여 실제 방향을 정량화하기 어려운 경우에 대한 모호성을 고려한 것으로, 예측 방향이 실측 결과와 동일하거나 이웃한 방향의 경우 정확한 것으로 간주될 때의 정확도를 의미한다.

[0087]

MCRoI 풀링 레이어를 포함하는 'Proposed' 네트워크는 객체 피처, 로컬 컨텍스트 피처, 글로벌 컨텍스트 피처, 즉 세 가지 종류의 피처를 모두 결합하여 미래의 움직임을 예측하기 때문에 전반적으로 기존 네트워크보다 성능이 우수함을 알 수 있다.

[0088]

방향, 속도, 행동을 클래스로 한 인스턴스 레벨의 움직임 예측은 딥러닝을 활용하는데 있어 픽셀 단위의 움직임 예측에 비해 다음의 이점이 있다: (1) 픽셀 단위의 어노테이션의 경우 시간과 비용이 많이 들어 DNN을 학습하기에 충분한 데이터를 확보하기 어려운 측면이 있다. (2) 객체 단위의 움직임 모델링은 인간이 움직임을 인식하고 이해하는 과정과 더욱 유사할 뿐만 아니라 기존의 검출(detection) 및 추적(tracking)과 같은 다양한 컴퓨터 비전 기술과 융합이 용이하다.

[0089]

본 발명의 실시예에서는 다양한 수준의 의미 정보를 통합하여 학습한 DNN를 이용하여 단일 정지 영상에서 객체의 미래 움직임을 정확하게 효율적으로 예측할 수 있고 미래 움직임에 대한 신뢰할 수 있는 예측 성능을 제공할 수 있다.

[0090]

상기한 미래 움직임 예측 알고리즘은 객체 트래킹을 더욱 효율적으로 만들 수 있다.

[0091]

연속된 프레임 간에 단일 객체는 물론이고 다중 객체에 대해서도 각 객체 별 미래 움직임 예측 결과를 이용하여 해당 객체의 트래킹을 위한 탐색 영역을 줄일 수 있다.

[0092]

미래 움직임 예측 결과를 사용하여 목표 객체의 탐색 영역을 줄이면 객체를 보다 신속하고 효율적으로 트래킹할 수 있다.

[0093]

도 8은 (a) 기존 트래킹 방식의 탐색 영역(81)과, (b) 미래 움직임 예측 결과를 활용한 트래킹 방식의 탐색 영역(801)을 비교한 것이다.

[0094]

기존에는 가우스 분포에서 기준선을 샘플링하여 사각 윈도우에 해당되는 탐색 영역(81) 내에서 검색 대상을 선

택하는 방식인 반면에, 본 발명에 따른 트래킹 방식에서는 미래 움직임 예측 결과를 바탕으로 탐색 영역(801)을 줄일 수 있다. 이때, 탐색 영역(801)은 미래 움직임으로 예측된 방향, 속도, 행동에 따라 결정될 수 있다. 일례로, 미래 움직임으로 예측된 속도가 빠를수록 탐색 영역(801)의 크기가 커지고 속도가 느릴수록 탐색 영역(801)의 크기가 작아질 수 있다. 예를 들어, 객체 트래킹부(230)는 미래 움직임의 속도가 '정지'일 때 기존 탐색 영역(81)이나 사전에 정해진 크기의 탐색 영역보다 4배 작은 탐색 영역(801)을 적용할 수 있다. 다른 예로는, 객체 트래킹부(230)는 미래 움직임의 방향에 대응되는 부채꼴 모양의 탐색 영역(801)을 결정할 수 있다.

[0095] 본 발명에서는 객체 트래킹을 위한 탐색 영역을 줄일 수 있어 기존보다 작은 영역을 탐색하여 트래킹 속도를 향상시킬 수 있다.

[0096] 낮은 프레임 레이트의 동영상에서 다중 객체 트래킹을 위해서는 추적별 탐지 방법(tracking-by-detection method)을 이용하여 정교한 트래킹을 수행할 수 있다. 일례로, 다중 객체 트래킹을 위해 각 객체의 미래 움직임 예측 결과를 바탕으로 탐색 영역을 줄이면서 아울러 샘플링 밀도를 높일 수 있다.

[0097] 객체의 미래 움직임은 객체 주변의 배경 정보와 관련하여 객체의 상대적인 움직임으로 예측되기 때문에 배경이 이동하는 경우 미래 움직임을 이용한 탐색 영역의 축소는 무효가 될 수 있다. 이에, 본 발명에서는 키포인트 매칭(key point matching)과 아핀(affine) 변환 등의 알고리즘을 사용하여 배경 이동을 우선 보정한 후 객체 트래킹을 수행할 수 있다.

[0098] 객체 트래킹 이외에도 미래 움직임 예측 알고리즘을 단일 이미지의 군중 분석에 적용하는 것 또한 가능하다.

[0099] 일례로, 움직임 예측부(220)는 군중의 각 보행자의 예측 방향을 사용하여 군중을 여러 클러스터로 분할한 후 각 클러스터의 그룹 방향을 예측할 수 있다. 이때, 클러스터링을 위해 K-평균(mean) 알고리즘을 사용한다. 두 인스턴스 간의 거리를 계산하려면 수학적 3의 가중 거리를 사용할 수 있다.

[0100] [수학적식 3]

$$D = D_{\text{Euc}} + \lambda D_{\text{FM}}$$

[0102] 여기서, D_{Euc} 는 인스턴스들 사이의 유클리드 거리이고, D_{FM} 은 방향적 지수의 주기적인 차이를 의미한다. 예를 들어, 인접한 방향 사이의 D_{FM} 은 1이고 반대 방향에서 최대 D_{FM} 은 4이다. 또한, λ 은 가중 파라미터를 의미한다.

[0103] 클러스터링 이후 각 클러스터의 그룹 방향을 획득하게 되고, 이를 위해 각 방향에 대해 클러스터 내의 인스턴스의 방향 확률의 합계를 계산할 수 있다. 이때, 확률 벡터는 도 6의 DNN(600)의 움직임 예측부(220)에 의해 생성될 수 있으며, 확률의 합이 최대인 방향을 해당 클러스터의 그룹 방향으로 결정할 수 있다.

[0104] 도 9의 (a)는 각 보행자의 이동 방향을 예측한 결과를 도시한 것이고, (b)는 클러스터링을 통해 클러스터링된 그룹의 이동 방향을 예측한 결과를 도시한 것이다. 도 9의 (b)에 도시한 이미지와 같이 데이터 클러스터링에 기초하여 보행자의 방향 정보를 압축하여 군중 분석 결과로서 제공할 수 있다.

[0105] 이처럼 본 발명의 실시예들에 따르면, 단일 정지 영상에서 객체의 미래 움직임을 예측할 수 있어 프레임 레이트가 현저히 낮은 동영상이나 프레임 드롭이 발생하는 경우에도 트래킹 성능 저하를 예방하고 안정적인 트래킹을 보장할 수 있다. 그리고, 본 발명의 실시예들에 따르면, 단일 정지 영상인 일 프레임에서 객체의 미래 움직임을 예측하여 예측 결과를 기반으로 다음 프레임에서 객체를 탐색함으로써 객체 탐색 영역을 효율적으로 줄일 수 있다.

[0106] 이상에서 설명된 장치는 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성요소는, 프로세서, 컨트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPGA(field programmable gate array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 어플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설

명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소 (processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 컨트롤러를 포함할 수 있다. 또한, 병렬 프로세서 (parallel processor)와 같은, 다른 처리 구성 (processing configuration)도 가능하다.

[0107] 소프트웨어는 컴퓨터 프로그램 (computer program), 코드 (code), 명령 (instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로 (collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소 (component), 물리적 장치, 컴퓨터 저장 매체 또는 장치에 구체화 (embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

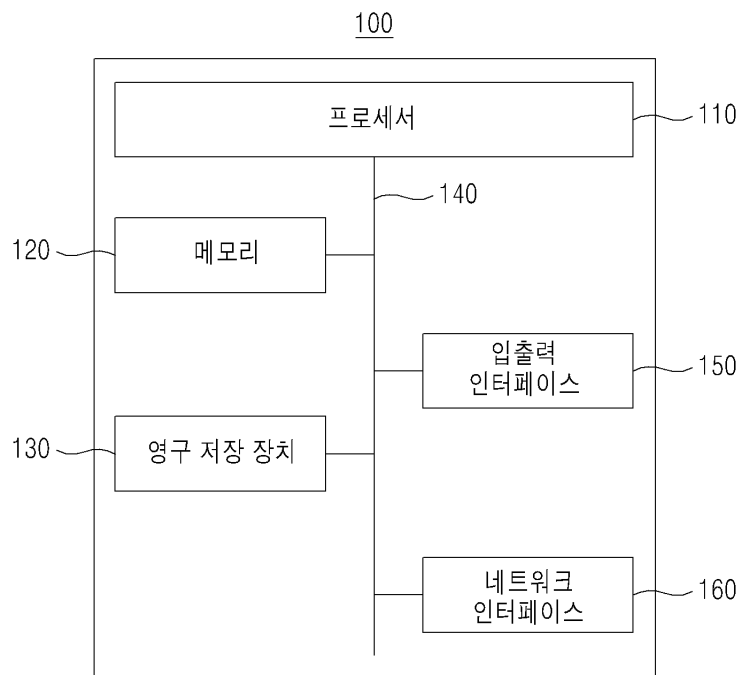
[0108] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 이때, 매체는 컴퓨터로 실행 가능한 프로그램을 계속 저장하거나, 실행 또는 다운로드를 위해 임시 저장하는 것일 수도 있다. 또한, 매체는 단일 또는 수 개의 하드웨어가 결합된 형태의 다양한 기록수단 또는 저장수단일 수 있는데, 어떤 컴퓨터 시스템에 직접 접속되는 매체에 한정되지 않고, 네트워크 상에 분산 존재하는 것일 수도 있다. 매체의 예시로는, 하드 디스크, 플로피 디스크 및 자기 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, 어플리케이션을 유통하는 앱 스토어나 기타 다양한 소프트웨어를 공급 내지 유통하는 사이트, 서버 등에서 관리하는 기록매체 내지 저장매체도 들 수 있다.

[0109] 이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

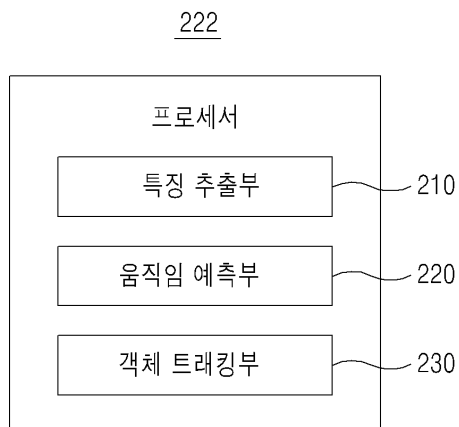
[0110] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

도면

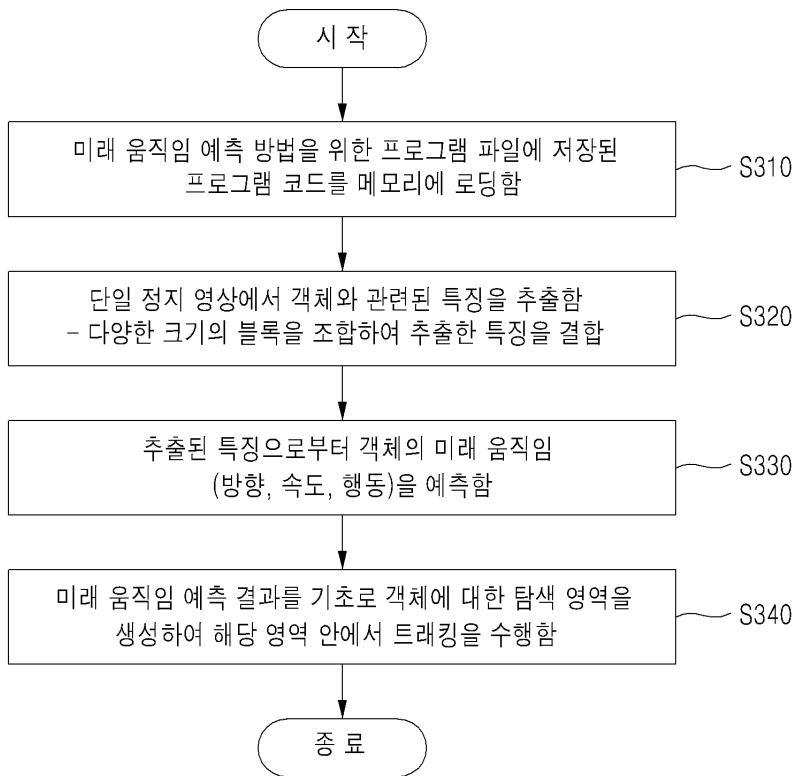
도면1



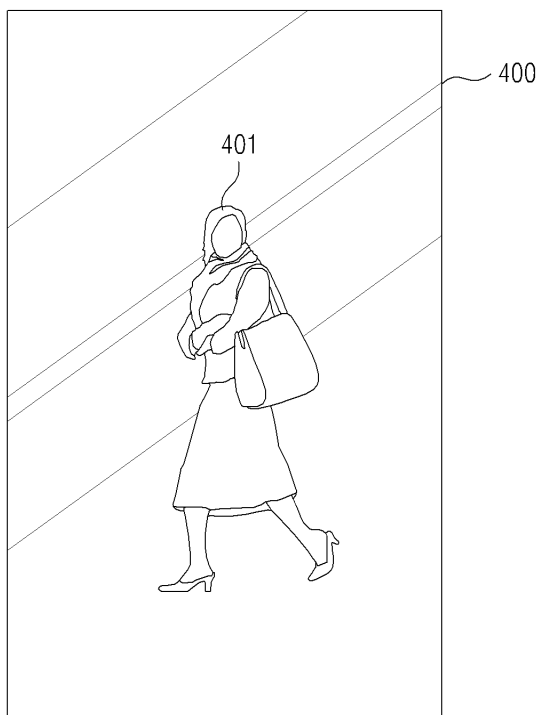
도면2



도면3

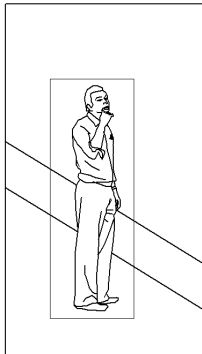


도면4

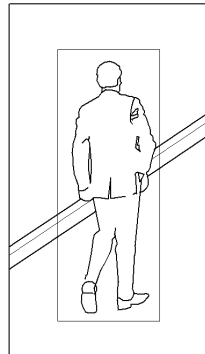


도면5

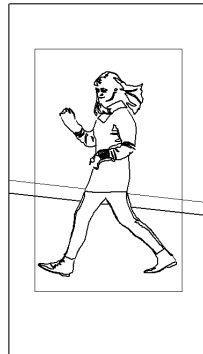
510



Stop

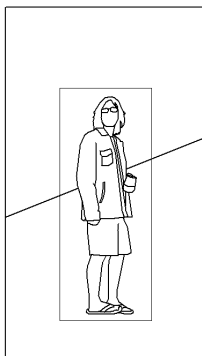


Slow

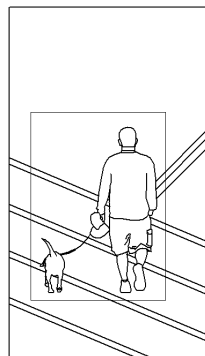


Fast

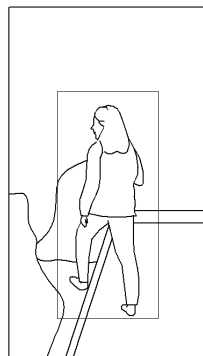
520



Sidewalk

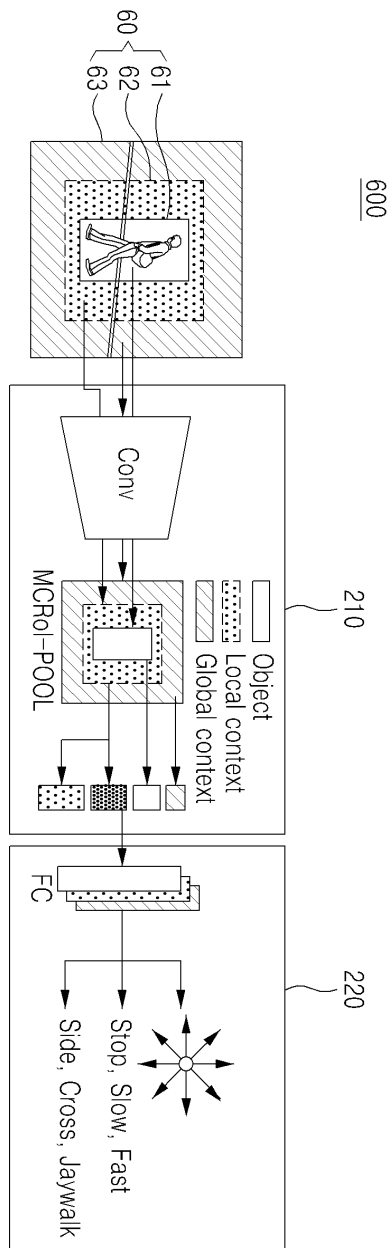


Crosswalk



Jaywalk

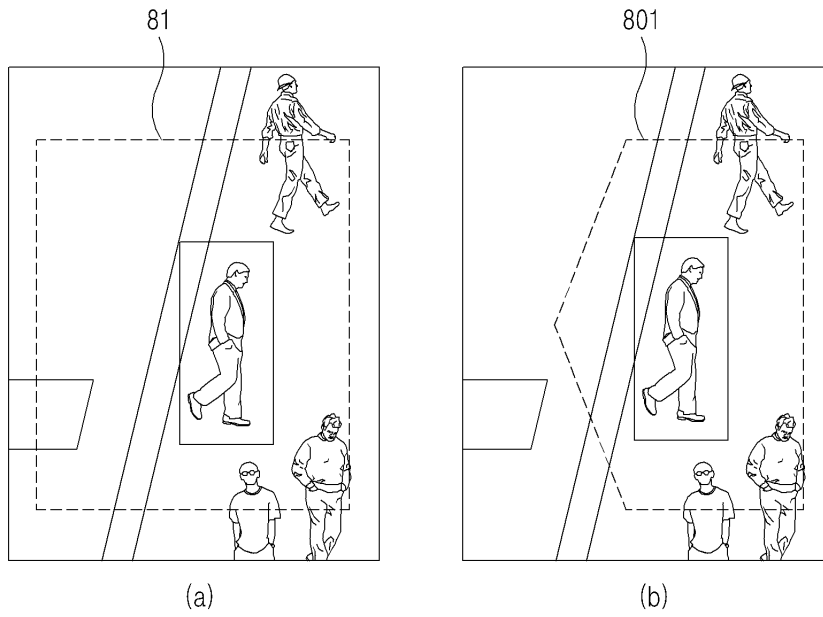
도면6



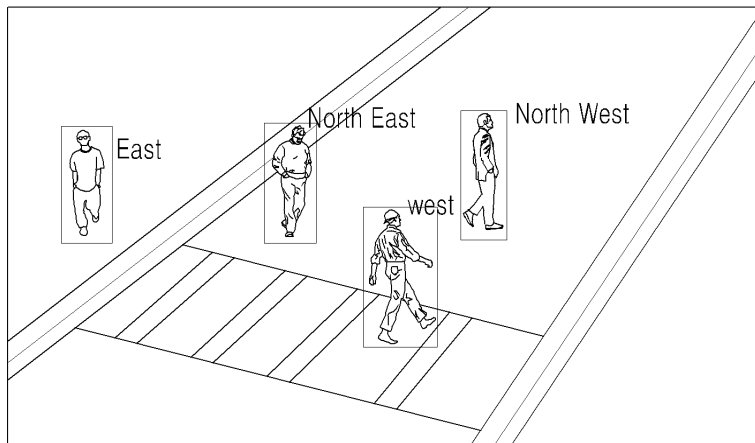
도면7

Network architecture	Direction	Direction+	Velocity	Action
Object	76.2	<u>95.3</u>	90.0	87.2
Local context	75.0	94.1	90.6	87.2
Global context	74.8	94.23	<u>90.4</u>	89.4
Object + Local	<u>76.8</u>	95.2	90.3	86.8
Proposed	77.2	95.7	89.0	<u>88.3</u>

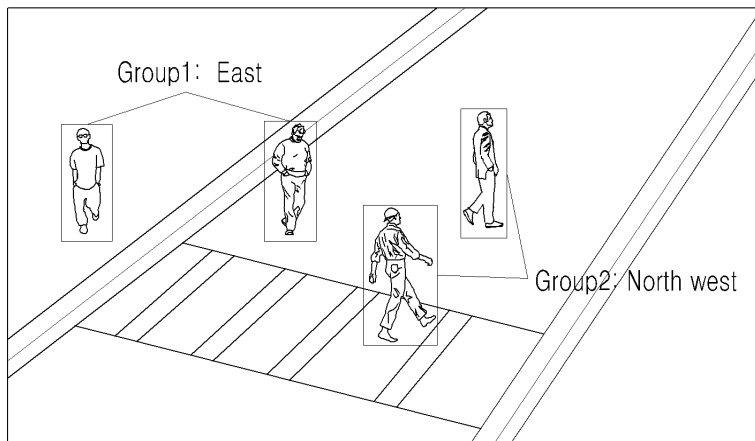
도면8



도면9



(a)



(b)