



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0126699
(43) 공개일자 2024년08월21일

(51) 국제특허분류(Int. Cl.)
G06V 20/64 (2022.01) G06V 10/40 (2022.01)
G06V 10/80 (2022.01)
(52) CPC특허분류
G06V 20/64 (2023.08)
G06V 10/40 (2023.08)
(21) 출원번호 10-2023-0019517
(22) 출원일자 2023년02월14일
심사청구일자 2023년02월14일

(71) 출원인
충남대학교산학협력단
대전광역시 유성구 대학로 99 (궁동, 충남대학교)
(72) 발명자
조동현
대전광역시 유성구 대학로 99 충남대학교 제 2공학관 407호
이전수
대전광역시 유성구 대학로 99 충남대학교 제 2공학관 472호
(74) 대리인
제일특허법인(유)

전체 청구항 수 : 총 16 항

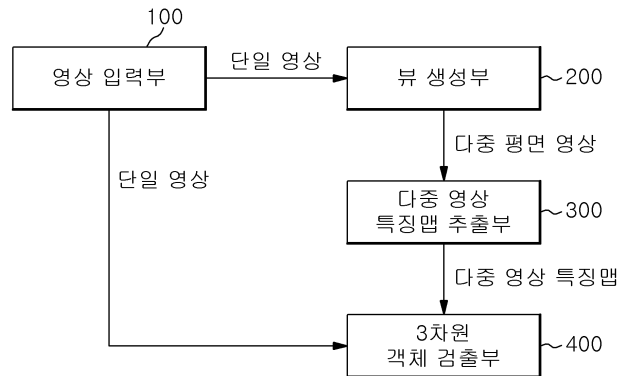
(54) 발명의 명칭 다중 평면 영상으로부터 3차원 모델링을 수행하기 위한 객체 검출 방법, 객체 검출 장치

(57) 요약

본 발명은 다중 평면 영상으로부터 3차원 모델링을 수행하기 위한 객체 검출 기술에 관한 것으로, 객체 검출 장치로 입력되는 단일 영상을 기초로 단일 영상 특징맵을 추출하고, 단일 영상으로부터 다중 평면 영상을 생성하며, 다중 평면 영상을 기초로 다중 영상 특징맵을 추출하고, 단일 영상 특징맵과 다중 영상 특징맵을 기초로 단일 영상의 3차원 객체를 검출할 수 있다.

대 표 도 - 도2

10



(52) CPC특허분류

G06V 10/806 (2023.08)

G06T 2207/10028 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711137685
과제번호	2021R1A4A1032580
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	기초연구실지원사업
연구과제명	협력 지능형 엣지를 위한 테라헤르츠 무선 시스템 연구실
기 여 율	1/1
과제수행기관명	충남대학교
연구기간	2021.06.01 ~ 2024.02.29

명세서

청구범위

청구항 1

다중 평면 영상으로부터 3차원 모델링을 수행하기 위한 객체 검출 장치의 객체 검출 방법에 있어서,
상기 객체 검출 장치로 입력되는 단일 영상을 기초로 단일 영상 특징맵을 추출하는 단계;
상기 단일 영상으로부터 다중 평면 영상을 생성하는 단계;
상기 다중 평면 영상을 기초로 다중 영상 특징맵을 추출하는 단계; 및
상기 단일 영상 특징맵과 상기 다중 영상 특징맵을 기초로 상기 단일 영상의 3차원 객체를 검출하는 단계;를 포함하는
객체 검출 방법.

청구항 2

제 1 항에 있어서,
상기 3차원 객체를 검출하는 단계는,
상기 단일 영상 특징맵과 상기 다중 영상 특징맵을 융합하여 융합된 특징맵을 생성하는 단계; 및
상기 융합된 특징맵을 기초로 상기 단일 영상의 3차원 객체를 검출하는 단계;를 포함하는
객체 검출 방법.

청구항 3

제 2 항에 있어서,
상기 융합된 특징맵은 시각 정보 및 깊이 정보를 포함하는
객체 검출 방법.

청구항 4

제 3 항에 있어서,
상기 3차원 객체를 검출하는 단계는,
상기 깊이 정보가 상기 시각 정보에 반영되어 상기 단일 영상의 3차원 객체를 검출하도록 학습되는 단계를 더 포함하는
객체 검출 방법.

청구항 5

제 4 항에 있어서,
상기 학습되는 단계는,
상기 3차원 객체의 검출 결과를 기초로 손실 함수를 계산하는 단계; 및

상기 손실 함수가 최소가 되도록 네트워크 파라미터를 갱신하는 단계;를 포함하는 객체 검출 방법.

청구항 6

제 1 항에 있어서,
상기 다중 평면 영상을 생성하는 단계는,
상기 단일 영상을 기초로 상기 다중 평면 영상의 평면의 개수에 대응하는 채널을 생성하는 단계; 및
상기 채널과 상기 단일 영상의 곱 연산을 통해 상기 다중 평면 영상을 생성하는 단계;를 포함하는 객체 검출 방법.

청구항 7

입력되는 단일 영상으로부터 다중 평면 영상을 생성하는 뷰 생성부;
상기 다중 평면 영상을 기초로 다중 영상 특징맵을 추출하는 다중 영상 특징맵 추출부; 및
상기 단일 영상으로부터 단일 영상 특징맵을 추출하고, 상기 단일 영상 특징맵과 상기 다중 영상 특징맵을 기초로 상기 단일 영상의 3차원 객체를 검출하는 3차원 객체 검출부;를 포함하는 객체 검출 장치.

청구항 8

제 7 항에 있어서,
상기 3차원 객체 검출부는,
상기 단일 영상 특징맵을 추출하는 단일 영상 특징맵 추출부;
상기 단일 영상 특징맵과 상기 다중 영상 특징맵을 융합하여 융합된 특징맵을 생성하는 특징맵 융합부; 및
상기 융합된 특징맵을 기초로 상기 단일 영상의 3차원 객체를 검출하는 객체 검출부;를 포함하는 객체 검출 장치.

청구항 9

제 8 항에 있어서,
상기 융합된 특징맵은 시각 정보 및 깊이 정보를 포함하는 객체 검출 장치.

청구항 10

제 9 항에 있어서,
상기 객체 검출부는,
상기 깊이 정보가 상기 시각 정보에 반영되어 상기 단일 영상의 3차원 객체를 검출하도록 학습되는

객체 검출 장치.

청구항 11

제 10 항에 있어서,
상기 객체 검출부는,
상기 3차원 객체의 검출 결과를 기초로 손실 함수를 계산하고,
상기 손실 함수가 최소가 되도록 네트워크 파라미터를 갱신하는
객체 검출 장치.

청구항 12

제 7 항에 있어서,
상기 뷰 생성부는,
상기 단일 영상을 기초로 상기 다중 평면 영상의 평면의 개수에 대응하는 채널을 생성하고,
상기 채널과 상기 단일 영상의 곱 연산을 통해 상기 다중 평면 영상을 생성하는
객체 검출 장치.

청구항 13

제 7 항에 있어서,
상기 다중 영상 특징맵 추출부는 딥러닝 네트워크의 3차원 컨벌루션 레이어(convolution layer)를 포함하는
객체 검출 장치.

청구항 14

제 8 항에 있어서,
상기 단일 영상 특징맵 추출부는 딥러닝 네트워크의 2차원 컨벌루션 레이어를 포함하는
객체 검출 장치.

청구항 15

컴퓨터 프로그램을 저장하고 있는 컴퓨터 판독 가능 기록매체로서,
상기 컴퓨터 프로그램은,
다중 평면 영상으로부터 3차원 모델링을 수행하기 위한 객체 검출 장치의 객체 검출 방법을 프로세서가 수행하도록 하기 위한 명령어를 포함하고,
상기 방법은,
상기 객체 검출 장치로 입력되는 단일 영상을 기초로 단일 영상 특징맵을 추출하는 단계;
상기 단일 영상으로부터 다중 평면 영상을 생성하는 단계;
상기 다중 평면 영상을 기초로 다중 영상 특징맵을 추출하는 단계; 및

상기 단일 영상 특징맵과 상기 다중 영상 특징맵을 기초로 상기 단일 영상의 3차원 객체를 검출하는 단계;를 포함하는

컴퓨터 판독 가능한 기록매체.

청구항 16

컴퓨터 판독 가능 기록매체에 저장된 컴퓨터 프로그램으로서,

상기 컴퓨터 프로그램은,

다중 평면 영상으로부터 3차원 모델링을 수행하기 위한 객체 검출 장치의 객체 검출 방법을 프로세서가 수행하도록 하기 위한 명령어를 포함하고,

상기 방법은,

상기 객체 검출 장치로 입력되는 단일 영상을 기초로 단일 영상 특징맵을 추출하는 단계;

상기 단일 영상으로부터 다중 평면 영상을 생성하는 단계;

상기 다중 평면 영상을 기초로 다중 영상 특징맵을 추출하는 단계; 및

상기 단일 영상 특징맵과 상기 다중 영상 특징맵을 기초로 상기 단일 영상의 3차원 객체를 검출하는 단계;를 포함하는

기록매체에 저장된 컴퓨터 프로그램.

발명의 설명

기술 분야

[0001] 본 발명은 3차원 객체 검출(3D object detection) 기술과 관련이 있다.

배경 기술

[0003] 3차원 객체 검출 기술은 입력 데이터 내부에 위치하고 있는 3차원 물체의 여러 가지 속성들(예를 들어, 크기, 위치, 회전각 등)을 추정하는 기술로서, 최근 자율주행 기술이 4차 산업혁명 핵심 기술 중 하나로 각광을 받으면서 3차원 객체 검출 기술에 대한 연구가 활발히 진행되고 있다.

[0004] 기존의 연구들은 객체 검출 과정에서 해당 장면의 단일 채널 깊이 맵(depth map)을 속성과 함께 추론하여 3차원 속성 추정에 도움을 주는 방향으로 진행되고 있다.

선행기술문헌

특허문헌

[0006] (특허문헌 0001) 공개특허공보 제10-2021-0153989호 (2021년12월20일 공개)

발명의 내용

해결하려는 과제

[0007] 본 발명의 실시예에서는, 단일 채널 깊이 맵을 이용하는 기존의 방법과는 다르게, 영상 합성(view synthesis) 과정에서 얻어진 고차원 정보인 다중 평면 영상(multi-plane image, MPI)를 이용하여 RGB(red, green, blue)

기반의 객체 탐지기의 성능을 향상시킬 수 있는 객체 검출 기술을 제안하고자 한다.

[0008] 본 발명이 해결하고자 하는 과제는 상기에서 언급한 것으로 제한되지 않으며, 언급되지 않은 또 다른 해결하고자 하는 과제는 아래의 기재들로부터 본 발명이 속하는 통상의 지식을 가진 자에 의해 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

[0010] 본 발명의 실시예에 따르면, 다중 평면 영상으로부터 3차원 모델링을 수행하기 위한 객체 검출 장치의 객체 검출 방법에 있어서, 상기 객체 검출 장치로 입력되는 단일 영상을 기초로 단일 영상 특징맵을 추출하는 단계; 상기 단일 영상으로부터 다중 평면 영상을 생성하는 단계; 상기 다중 평면 영상을 기초로 다중 영상 특징맵을 추출하는 단계; 및 상기 단일 영상 특징맵과 상기 다중 영상 특징맵을 기초로 상기 단일 영상의 3차원 객체를 검출하는 단계;를 포함하는 객체 검출 방법을 제공할 수 있다.

[0011] 여기서, 상기 3차원 객체를 검출하는 단계는, 상기 단일 영상 특징맵과 상기 다중 영상 특징맵을 융합하여 융합된 특징맵을 생성하는 단계; 및 상기 융합된 특징맵을 기초로 상기 단일 영상의 3차원 객체를 검출하는 단계;를 포함할 수 있다.

[0012] 또한, 상기 융합된 특징맵은 시각 정보 및 깊이 정보를 포함할 수 있다.

[0013] 또한, 상기 3차원 객체를 검출하는 단계는, 상기 깊이 정보가 상기 시각 정보에 반영되어 상기 단일 영상의 3차원 객체를 검출하도록 학습되는 단계를 더 포함할 수 있다.

[0014] 또한, 상기 학습되는 단계는, 상기 3차원 객체의 검출 결과를 기초로 손실 함수를 계산하는 단계; 및 상기 손실 함수가 최소가 되도록 네트워크 파라미터를 갱신하는 단계;를 포함할 수 있다.

[0015] 또한, 상기 다중 평면 영상을 생성하는 단계는, 상기 단일 영상을 기초로 상기 다중 평면 영상의 평면의 개수에 대응하는 채널을 생성하는 단계; 및 상기 채널과 상기 단일 영상의 곱 연산을 통해 상기 다중 평면 영상을 생성하는 단계;를 포함할 수 있다.

[0016] 본 발명의 실시예에 따르면, 입력되는 단일 영상으로부터 다중 평면 영상을 생성하는 뷰 생성부; 상기 다중 평면 영상을 기초로 다중 영상 특징맵을 추출하는 다중 영상 특징맵 추출부; 및 상기 단일 영상으로부터 단일 영상 특징맵을 추출하고, 상기 단일 영상 특징맵과 상기 다중 영상 특징맵을 기초로 상기 단일 영상의 3차원 객체를 검출하는 3차원 객체 검출부;를 포함하는 객체 검출 장치를 제공할 수 있다.

[0017] 여기서, 상기 3차원 객체 검출부는, 상기 단일 영상 특징맵을 추출하는 단일 영상 특징맵 추출부; 상기 단일 영상 특징맵과 상기 다중 영상 특징맵을 융합하여 융합된 특징맵을 생성하는 특징맵 융합부; 및 상기 융합된 특징맵을 기초로 상기 단일 영상의 3차원 객체를 검출하는 객체 검출부;를 포함할 수 있다.

[0018] 또한, 상기 융합된 특징맵은 시각 정보 및 깊이 정보를 포함할 수 있다.

[0019] 또한, 상기 객체 검출부는, 상기 깊이 정보가 상기 시각 정보에 반영되어 상기 단일 영상의 3차원 객체를 검출하도록 학습될 수 있다.

[0020] 또한, 상기 객체 검출부는, 상기 3차원 객체의 검출 결과를 기초로 손실 함수를 계산하고, 상기 손실 함수가 최소가 되도록 네트워크 파라미터를 갱신할 수 있다.

[0021] 또한, 상기 뷰 생성부는, 상기 단일 영상을 기초로 상기 다중 평면 영상의 평면의 개수에 대응하는 채널을 생성하고, 상기 채널과 상기 단일 영상의 곱 연산을 통해 상기 다중 평면 영상을 생성할 수 있다.

[0022] 또한, 상기 다중 영상 특징맵 추출부는 딥러닝 네트워크의 3차원 컨벌루션 레이어(convolution layer)를 포함할 수 있다.

[0023] 또한, 상기 단일 영상 특징맵 추출부는 딥러닝 네트워크의 2차원 컨벌루션 레이어를 포함할 수 있다.

[0024] 본 발명의 실시예에 따르면, 컴퓨터 프로그램을 저장하고 있는 컴퓨터 판독 가능 기록매체로서, 상기 컴퓨터 프로그램은, 다중 평면 영상으로부터 3차원 모델링을 수행하기 위한 객체 검출 장치의 객체 검출 방법을 프로세서가 수행하도록 하기 위한 명령어를 포함하고, 상기 방법은, 상기 객체 검출 장치로 입력되는 단일 영상을 기초로 단일 영상 특징맵을 추출하는 단계; 상기 단일 영상으로부터 다중 평면 영상을 생성하는 단계; 상기 다중 평면

면 영상을 기초로 다중 영상 특징맵을 추출하는 단계; 및 상기 단일 영상 특징맵과 상기 다중 영상 특징맵을 기초로 상기 단일 영상의 3차원 객체를 검출하는 단계;를 포함할 수 있다.

- [0025] 본 발명의 실시예에 따르면, 컴퓨터 관독 가능 기록매체에 저장된 컴퓨터 프로그램으로서, 상기 컴퓨터 프로그램은, 다중 평면 영상으로부터 3차원 모델링을 수행하기 위한 객체 검출 장치의 객체 검출 방법을 프로세서가 수행하도록 하기 위한 명령어를 포함하고, 상기 방법은, 상기 객체 검출 장치로 입력되는 단일 영상을 기초로 단일 영상 특징맵을 추출하는 단계; 상기 단일 영상으로부터 다중 평면 영상을 생성하는 단계; 상기 다중 평면 영상을 기초로 다중 영상 특징맵을 추출하는 단계; 및 상기 단일 영상 특징맵과 상기 다중 영상 특징맵을 기초로 상기 단일 영상의 3차원 객체를 검출하는 단계;를 포함할 수 있다.

발명의 효과

- [0027] 본 발명의 실시예에 의하면, 다중 평면 영상을 이용한 3차원 객체 탐지 네트워크의 학습 기법을 통해 기존의 RGB 영상 기반의 탐지 기법의 한계점을 극복하면서 3차원 객체 검출 성능을 개선시킬 수 있다. 또한, 본 발명의 실시예에 의하면, 일반적인 깊이 맵보다 고차원의 정보를 제공할 수 있으며, 기존 레거시(legacy) 객체 탐지 기술에 대한 응용 활용도와 범용성을 높일 수 있다.

도면의 간단한 설명

- [0029] 도 1은 3차원 객체 검출 과정에서 추정해야 하는 7개의 파라미터들을 예시적으로 표현한 도면이다.
- 도 2는 본 발명의 실시예에 따른 다중 평면 영상으로부터 3차원 모델링을 수행하기 위한 객체 검출 장치(10)의 블록도이다.
- 도 3은 도 1의 객체 검출 장치(10)의 뷰 생성부(200)를 통해 생성되는 다중 평면 영상과 각 채널을 예시적으로 나타낸 도면이다.
- 도 4는 도 1의 객체 검출 장치(10)의 3차원 객체 검출부(400)에 대한 구체적인 기능 블록도이다.
- 도 5는 본 발명의 실시예에 따른 다중 평면 영상으로부터 3차원 모델링을 수행하기 위한 객체 검출 장치(10)의 객체 검출 방법을 예시적으로 설명하는 흐름도이다.
- 도 6은 기존의 단일 영상 특징 맵으로부터 객체를 검출하는 과정을 예시적으로 설명하는 흐름도이다.

발명을 실시하기 위한 구체적인 내용

- [0030] 본 발명의 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 상세하게 후술되어 있는 실시예들을 참조하면 명확해질 것이다. 그러나 본 발명은 이하에서 개시되는 실시예들에 한정되는 것이 아니라 다양한 형태로 구현될 수 있으며, 단지 본 실시예들은 본 발명의 개시가 완전하도록 하고, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에게 발명의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 발명의 범주는 청구항에 의해 정의될 뿐이다.
- [0031] 본 발명의 실시예들을 설명함에 있어서 공지 기능 또는 구성에 대한 구체적인 설명은 본 발명의 실시예들을 설명함에 있어 실제로 필요한 경우 외에는 생략될 것이다. 그리고 후술되는 용어들은 본 발명의 실시예에서의 기능을 고려하여 정의된 용어들로서 이는 사용자, 운용자의 의도 또는 관례 등에 따라 달라질 수 있다. 그러므로 그 정의는 본 명세서 전반에 걸친 내용을 토대로 내려져야 할 것이다.
- [0032] 3차원 객체 탐지를 위한 입력 데이터로는 일반적인 2차원 RGB 영상이나 LiDAR(light detection and ranging)와 같은 능동 센서를 통해 얻어진 3차원 포인트 클라우드(point cloud)를 활용할 수 있다.
- [0033] 도 1은 3차원 객체 탐지 네트워크에서 추정해야 하는 7개의 파라미터들을 예시적으로 나타내고 있다. 7개의 파라미터들은, 예컨대 3차원 포인트 클라우드에 대한 x축, y축 및 z축에 대한 좌표(x_c , y_c , z_c), 3차원 객체의 높이(h), 너비(w) 및 깊이(l), 3차원 객체 회전 각도(θ) 등을 포함할 수 있다.
- [0034] 정확한 위치 정보를 제공하는 3차원 포인트 클라우드를 활용하는 객체 탐지 네트워크의 경우, 탐지 난이도가 높은 물체에 대해서도 80% 이상의 높은 정확도를 보여줌에도 불구하고, 카메라에 비해 상대적으로 값이 비싸고 데

이터 저장 및 전처리가 까다로운 LiDAR와 같은 능동 센서를 활용해야 한다는 한계점이 존재한다.

- [0035] 이와는 달리, 차량용 카메라는 값이 저렴하여 이용하기에 부담이 적으나 카메라로부터 얻어진 2차원 RGB 영상은 이미 3차원 공간의 정보가 2차원 이미지 평면에 투영(projection)된 결과로 깊이에 대한 정보 손실이 발생한 상태이기 때문에, RGB 영상으로부터 물체의 3차원 속성을 추정하는 것은 유일한 해가 존재하지 않는 불량조건 문제(ill-posed problem)로서, 해법이 까다로우며 성능 자체도 탐지 난이도가 높은 물체에 대해 20% 미만의 정확도를 보여주는 수준에 머무르고 있다. 따라서, 3차원 포인트 클라우드 기반의 네트워크와는 큰 성능 차이가 존재한다.
- [0036] 이러한 한계점을 극복하기 위해 기존의 많은 방법들은, 객체 탐지 과정에서 해당 장면의 단일 채널 깊이 맵(depth map)까지 속성을 함께 추론하여 3차원 속성 추정에 도움을 주는 방향으로 연구가 진행되고 있다.
- [0037] 아울러, 컴퓨터 비전 분야에서 여러 시점에서 촬영한 이미지를 활용하여 새로운 시점의 영상을 생성하는 영상 합성(view synthesis)에 대한 연구가 활발히 진행되고 있다. 다른 시점에서 바라본 새로운 영상을 생성하기 위해서는 단순히 기존의 깊이 추정기(depth estimator)에서와 같이 이미지를 구성하는 각 픽셀마다의 깊이 값을 추정하는 것이 아니라 원하는 시작 시점과 도착 시점이 주어졌을 때 그 사이의 경로를 광선이 이동하는 과정에서 충돌이 발생할 확률을 추정하는 등 좀 더 수준 높은 이해가 필요하다.
- [0038] 이에 본 발명의 실시예에서는, 다중 평면 영상을 이용한 3차원 객체 탐지 네트워크의 학습 기법을 통해 기존의 RGB 영상 기반의 탐지 기법의 한계점을 극복하면서 3차원 객체 검출 성능을 개선할 수 있는 객체 탐지 기술을 제안하고자 한다.
- [0039] 이하, 첨부된 도면을 참조하여 본 발명의 실시예에 대해 상세히 설명하기로 한다.
- [0040] 도 2는 본 발명의 실시예에 따른 다중 평면 영상으로부터 3차원 모델링을 수행하기 위한 객체 검출 장치(10)의 블록도이다.
- [0041] 도 2에 도시한 바와 같이, 다중 평면 영상으로부터 3차원 모델링을 수행하기 위한 객체 검출 장치(10)는, 영상 입력부(100), 뷰 생성부(200), 다중 영상 특징맵 추출부(300) 및 3차원 객체 검출부(400)를 포함할 수 있다.
- [0042] 영상 입력부(100)는 객체 검출 장치(10)의 3차원 모델링을 위한 영상을 획득할 수 있다. 영상 입력부(100)에서 획득되는 영상은, 예를 들어 단일 RGB 영상을 포함할 수 있다.
- [0043] 뷰 생성부(200)는 영상 입력부(100)에서 획득되는 단일 영상을 기초로 다중 평면 영상을 생성할 수 있다.
- [0044] 뷰 생성부(200)는 네트워크 파라미터 θ_c 에 의해 정의되는 신경망으로서, 단일 RGB 영상을 입력으로 받아 사전에 정의된 하이퍼-파라미터인 평면의 개수 N 만큼의 채널을 갖는 다중 평면 영상을 생성할 수 있다.
- [0045] 단일 RGB 영상 x 와 다중 평면 영상 y_{alpha} 는 각각 독립적인 이미지를 의미하며, 다음 [수학식 1] 및 [수학식 2]로 표현될 수 있다.

수학식 1

[0046]
$$x \in \mathbb{R}^{3 \times H \times W}$$

수학식 2

[0047]
$$y_{alpha} \in \mathbb{R}^{N \times H \times W}$$

- [0048] 이 데이터의 각 채널은 원본 RGB 영상 x 의 곱 연산을 통해 아래 [수학식 3]과 같은 α 값으로 변환될 수 있다.

수학식 3

[0049] $y = (y_{alpha} \times x) \in R^{3N \times H \times W}$

[0050] [수학식 3]의 y 값은 다중 영상 특징맵 추출부(300)로 전달될 수 있다.

[0051] [수학식 1] 내지 [수학식 3]에서 H는 이미지의 높이, W는 이미지의 너비, N은 공간을 몇 개의 평면으로 나눌 것인지에 대한 파라미터를 각각 의미한다.

[0052] 도 3은 다중 평면 영상인 $y_{alpha} \in R^{N \times H \times W}$ 와 $y \in R^{3N \times H \times W}$ 로부터 추출된 몇 개의 채널($N=0, 10, 20, 30, 40, 50$)에 대한 예시를 각각 보여준다. 뷰 생성부(200)는 딥러닝 네트워크의 사전 학습 모델을 활용할 수 있으며, 평면의 개수를 결정하는 파라미터인 N은 데이터의 용량을 고려하여 32로 설정할 수 있다.

[0053] 다중 영상 특징맵 추출부(300)는 뷰 생성부(200)에서 생성된 다중 평면 영상을 기초로 다중 영상 특징맵을 추출할 수 있다.

[0054] 이러한 다중 영상 특징맵 추출부(300)는 네트워크 파라미터 θ_k 에 의해 정의되는 신경망으로서, 뷰 생성부(200)에서 생성된 $y \in R^{3N \times H \times W}$ 로부터 아래 [수학식 4]와 같은 다중 평면 영상에 대한 특징맵을 추출할 수 있다.

수학식 4

[0055] $f_{msi} \in R^{C \times H_i \times W_i}$

[0056] [수학식 4]에서 $H_f = \frac{1}{4}H$, $W_f = \frac{1}{4}W$ 를 각각 만족하도록 설정될 수 있다. [수학식 4]에서 1/4은 특징 맵의 너비(W)와 높이(H)가 4배 축소되어 출력된 것을 의미한다. 이러한 출력은, 예를 들어 맥스 풀링(max-pooling) 연산 기법을 통해 구현될 수 있다.

[0057] 이러한 다중 영상 특징맵 추출부(300)는 3차원 컨벌루션 레이어(convolution layer)로 구성된 ResNet 아키텍처를 이용하여 다중 평면 영상 $y = (y_{alpha} \times x) \in R^{3N \times H \times W}$ 를 구성하는 여러 개의 채널을 3차원 볼륨(volume)으로 해석하여 특징맵을 추출할 수 있다.

[0058] 3차원 객체 검출부(400)는 영상 입력부(100)에서 획득된 단일 영상을 기초로 단일 영상 특징맵을 추출하고, 추출된 단일 영상 특징맵과 다중 영상 특징맵 추출부(300)에서 추출된 다중 영상 특징맵을 기초로, 단일 영상의 3차원 객체를 검출할 수 있다.

[0059] 도 4는 도 1의 객체 검출 장치(10)의 3차원 객체 검출부(400)에 대한 구체적인 기능 블록도이다.

[0060] 3차원 객체 검출부(400)는, 영상 입력부(100)의 단일 영상으로부터 단일 영상 특징맵을 추출하는 단일 영상 특징맵 추출부(410), 단일 영상 특징맵 추출부(410)의 단일 영상 특징맵과 다중 영상 특징맵 추출부(300)의 다중 영상 특징맵을 융합하여 융합된 특징맵을 생성하는 특징맵 융합부(420), 특징맵 융합부(420)에 의해 융합된 특징맵을 기초로 단일 영상의 3차원 객체를 검출하는 객체 검출부(430)를 포함할 수 있다.

[0061] 단일 영상 특징맵 추출부(410)는 백본(backbone) 네트워크를 포함할 수 있으며, 영상 입력부(100)를 통해 입력되는 단일 RGB 영상 $x \in R^{3 \times H \times W}$ 를 기초로, 단일 RGB 영상에 대한 특징맵을 추출할 수 있다. 단일 영상 특징맵 추출부(410)에서 추출되는 특징맵은 아래 [수학식 5]와 같이 표현될 수 있다.

수학식 5

[0062] $f_{rob} \in R^{C \times H_i \times W_i}$

[0063] [수학식 5]에서, $H_f = \frac{1}{4}H$ 와 $W_f = \frac{1}{4}W$ 를 각각 만족하도록 설정될 수 있으며, 단일 영상 특징맵 추출부(410)를 구성하는 백본 네트워크는, 예를 들어 2차원 컨볼루션 레이어를 포함할 수 있다.

[0064] 특징맵 융합부(420)는 단일 영상 특징맵 추출부(410)에서 추출된 단일 영상 특징맵 $f_{rob} \in R^{C \times H_r \times W_r}$ 와, 다중 영상 특징맵 추출부(300)로부터 추출된 다중 영상 특징맵 $f_{mvi} \in R^{C \times H_r \times W_r}$ 를 융합하여 융합된 특징맵 f_{fused} 를 다음 [수학식 6]과 같이 생성할 수 있다.

수학식 6

[0065] $f_{fused} \in R^{C \times H_r \times W_r}$

[0066] 융합된 특징맵 f_{fused} 는 시각적인 정보뿐만 아니라 깊이에 대한 정보를 함께 포함할 수 있다.

[0067] 특징맵 융합부(420)의 융합은 단일 영상 특징맵과 다중 영상 특징맵 간의 덧셈 연산을 통해 진행될 수 있으며, 이는 아래 [수학식 7]로 표현될 수 있다.

수학식 7

[0068] $f_{fused} = f_{rob} + f_{mvi}$

[0069] 객체 검출부(430)는 특징맵 융합부(420)에서 얻어진 융합된 특징맵 f_{fused} 또는 단일 영상 특징맵 추출부(410)로부터 추출된 단일 영상 특징맵 $f_{rob} \in R^{C \times H_r \times W_r}$ 를 입력으로 받아 객체 검출 결과인 f_{det} 을 출력할 수 있다. 이때, 객체 검출부(430)는 객체 검출 결과 f_{det} 를 기반으로 손실이 계산되어 학습 과정에서 네트워크의 파라미터 θ_D 가 손실이 적어지는 방향으로 갱신될 수 있다.

[0070] 도 5는 본 발명의 실시예에 따른 다중 평면 영상으로부터 3차원 모델링을 수행하기 위한 객체 검출 장치(10)의 객체 검출 방법을 예시적으로 설명하는 흐름도이다.

[0071] 도 5에 도시한 바와 같이, 객체 검출 장치(10)는 영상 입력부(100)를 통해 입력되는 단일 영상을 기초로 단일 영상 특징맵을 추출할 수 있다(S102). 이때, 단일 영상 특징맵은 단일 영상 특징맵 추출부(410)를 통해 추출될 수 있으며, 단일 영상 특징맵 추출부(410)의 가중치만을 추출하고, 갱신이 되지 않도록 임시 저장될 수 있다(S104).

[0072] 이후, 객체 검출 장치(10)는 뷰 생성부(200)를 통해 영상 입력부(100)의 단일 영상으로부터 다중 평면 영상을 생성할 수 있다(S106).

[0073] 생성된 다중 평면 영상은 객체 검출 장치(10)의 다중 영상 특징맵 추출부(300)로 제공될 수 있으며, 다중 영상 특징맵 추출부(300)는 뷰 생성부(200)의 다중 평면 영상을 기초로 다중 영상 특징맵을 추출할 수 있다(S108).

[0074] 이후, 객체 검출 장치(10)는 3차원 객체 검출부(400) 내의 특징맵 융합부(420)를 통해 단일 영상 특징맵 추출부(410)의 단일 영상 특징맵과 다중 영상 특징맵 추출부(300)의 다중 영상 특징맵을 융합하여 융합된 특징맵을 생성할 수 있다(S110).

[0075] 융합된 특징맵이 생성되면, 객체 검출 장치(10)는 3차원 객체 검출부(400) 내의 객체 검출부(430)를 통해, 융합된 특징맵을 기초로 단일 영상의 3차원 객체를 검출할 수 있다(S112).

[0076] 한편, 도 6은 기존의 단일 영상 특징 맵으로부터 객체를 검출하는 과정을 예시적으로 설명하는 흐름도이다.

[0077] 도 6에 도시한 바와 같이, 객체 검출 장치(10)의 영상 입력부(100)로 단일 RGB 영상이 입력되면(S200), 객체 검출 장치(10)는 3차원 객체 검출부(400)의 단일 영상 특징맵 추출부(410)를 통해 단일 RGB 영상으로부터 단일 영상 특징맵을 추출할 수 있다(S202).

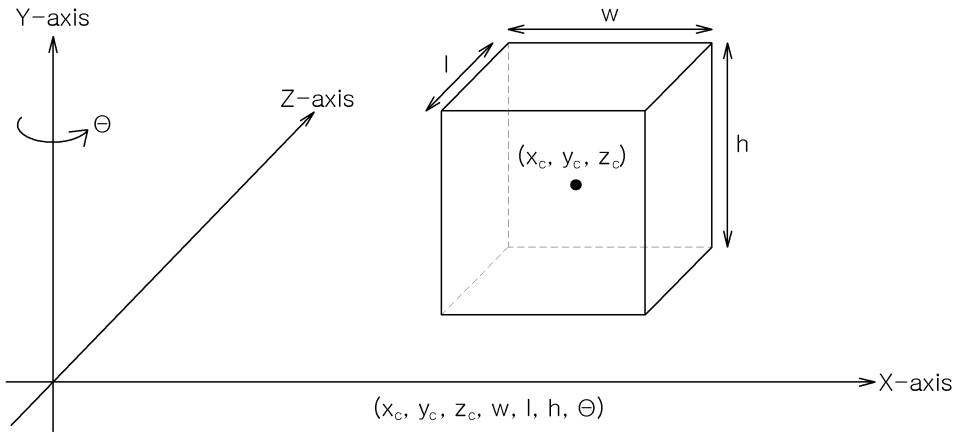
- [0078] 단일 영상 특징맵 추출부(410)에서 추출된 단일 영상 특징맵은 객체 검출부(430)로 제공될 수 있으며, 객체 검출부(430)는 단일 영상 특징맵을 기초로 학습을 수행하여 단일 영상에 대한 3차원 객체를 검출할 수 있다 (S204).
- [0079] 즉, 도 6의 과정은 다중 평면 영상을 이용하지 않고 단순 RGB 입력 $x \in \mathbb{R}^3 \times H \times W$ 만을 이용하여 3차원 객체 검출부(400)를 학습하는 것이며, 이 과정에서는 뷰 생성부(200)와 다중 영상 특징맵 추출부(300)의 학습 과정은 활용되지 않는다. 따라서, $f_{fused} = f_{rot}$ 조건을 만족할 수 있다.
- [0080] 이상 설명한 바와 같은 본 발명의 실시예에 의하면, 다중 평면 영상을 이용한 3차원 객체 탐지 네트워크의 학습 기법을 통해 기존의 RGB 영상 기반의 탐지 기법의 한계점을 극복하면서 3차원 객체 검출 성능을 개선할 수 있는 객체 탐지 기술을 제안함으로써, 일반적인 깊이 맵보다 고차원의 정보를 제공할 수 있으며, 기존 레거시 객체 탐지 기술에 대한 응용 활용도와 범용성을 높일 수 있을 것으로 기대된다.
- [0081] 한편, 첨부된 블록도의 각 블록과 흐름도의 각 단계의 조합들은 컴퓨터 프로그램 인스트럭션들에 의해 수행될 수도 있다. 이들 컴퓨터 프로그램 인스트럭션들은 범용 컴퓨터, 특수용 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비의 프로세서에 탑재될 수 있으므로, 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비의 프로세서를 통해 수행되는 그 인스트럭션들이 블록도의 각 블록에서 설명된 기능들을 수행하는 수단을 생성하게 된다.
- [0082] 이들 컴퓨터 프로그램 인스트럭션들은 특정 방식으로 기능을 구현하기 위해 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비를 지향할 수 있는 컴퓨터 이용 가능 또는 컴퓨터 판독 가능 기록매체(또는 메모리) 등에 저장되는 것도 가능하므로, 그 컴퓨터 이용 가능 또는 컴퓨터 판독 가능 기록매체(또는 메모리)에 저장된 인스트럭션들은 블록도의 각 블록에서 설명된 기능을 수행하는 인스트럭션 수단을 내포하는 제조 품목을 생산하는 것도 가능하다.
- [0083] 그리고, 컴퓨터 프로그램 인스트럭션들은 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비 상에 탑재되는 것도 가능하므로, 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비 상에서 일련의 동작 단계들이 수행되어 컴퓨터로 실행되는 프로세스를 생성해서 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비를 수행하는 인스트럭션들은 블록도의 각 블록에서 설명된 기능들을 실행하기 위한 단계들을 제공하는 것도 가능하다.
- [0084] 또한, 각 블록은 특정된 논리적 기능(들)을 실행하기 위한 적어도 하나 이상의 실행 가능한 인스트럭션들을 포함하는 모듈, 세그먼트 또는 코드의 일부를 나타낼 수 있다. 또, 몇 가지 대체 실시 예들에서는 블록들에서 언급된 기능들이 순서를 벗어나서 발생하는 것도 가능함을 주목해야 한다. 예컨대, 잇달아 도시되어 있는 두 개의 블록들은 사실 실질적으로 동시에 수행되는 것도 가능하고 또는 그 블록들이 때때로 해당하는 기능에 따라 역순으로 수행되는 것도 가능하다.

부호의 설명

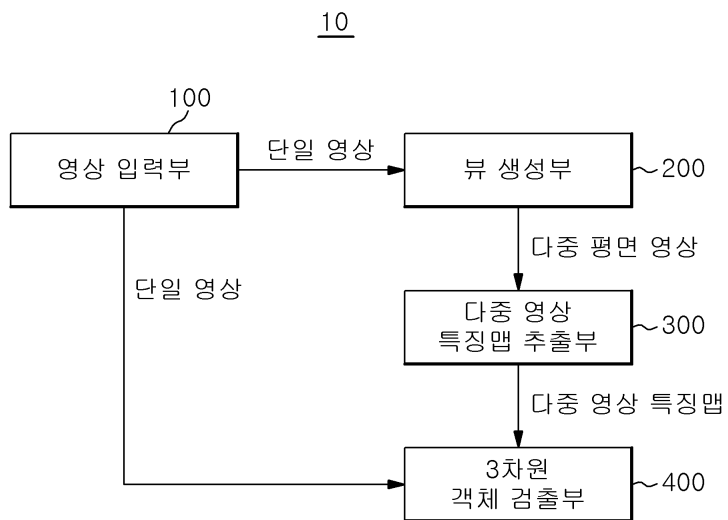
- [0086] 10: 객체 검출 장치
 100: 영상 입력부
 200: 뷰 생성부
 300: 다중 영상 특징맵 추출부
 400: 3차원 객체 검출부
 410: 단일 영상 특징맵 추출부
 420: 특징맵 융합부
 430: 객체 검출부

도면

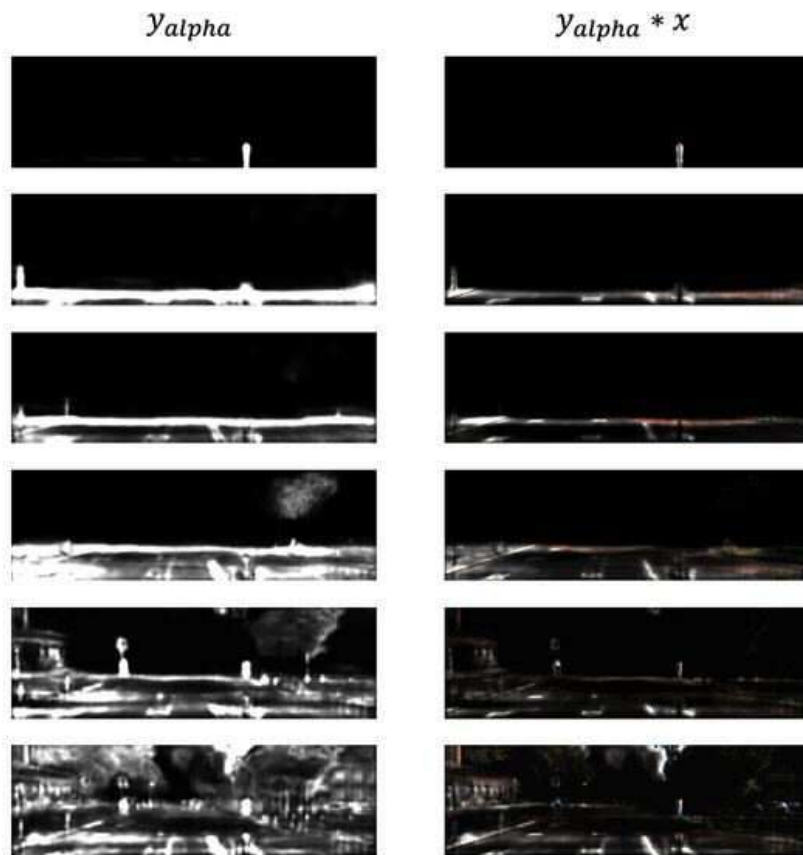
도면1



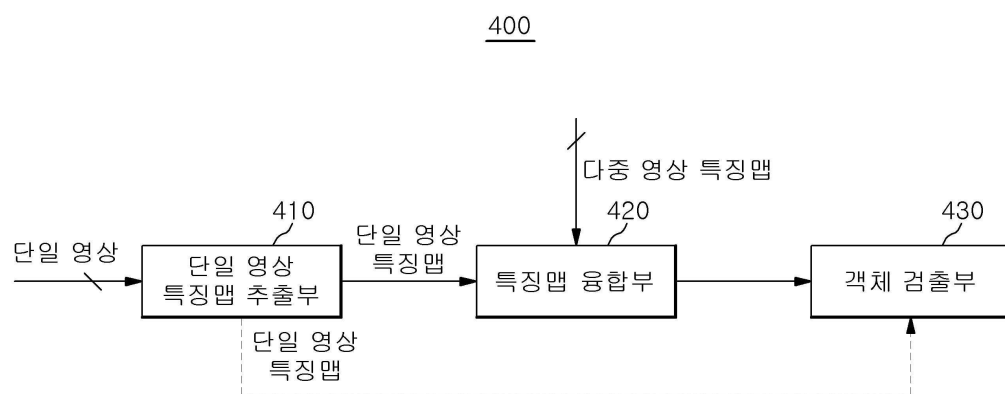
도면2



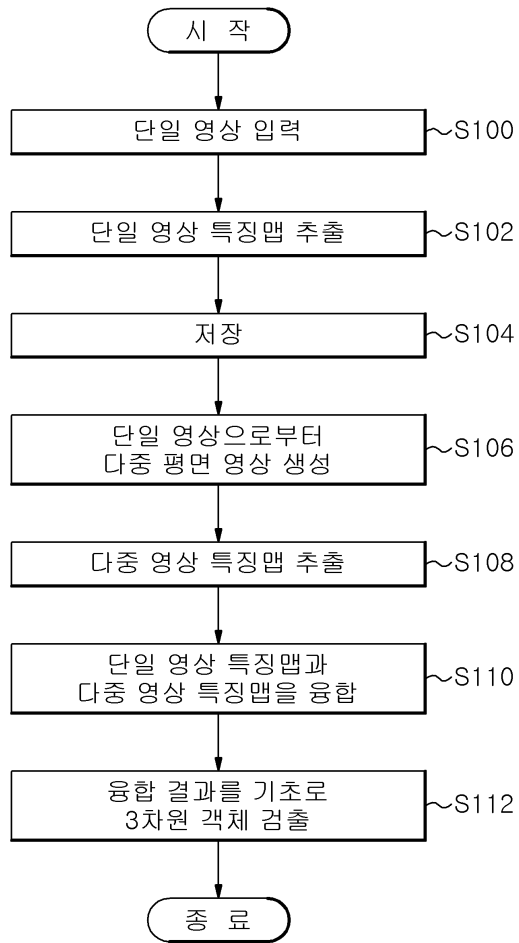
도면3



도면4



도면5



도면6

