



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0049328
(43) 공개일자 2023년04월13일

(51) 국제특허분류(Int. Cl.)

G06T 7/50 (2017.01) G01S 17/89 (2020.01)

G06N 3/04 (2023.01) G06N 3/08 (2023.01)

(52) CPC특허분류

G06T 7/50 (2017.01)

G01S 17/89 (2022.01)

(21) 출원번호 10-2021-0132328

(22) 출원일자 2021년10월06일

심사청구일자 2021년10월06일

(71) 출원인

세종대학교산학협력단

서울특별시 광진구 능동로 209 (군자동, 세종대학교)

(72) 발명자

최유경

서울특별시 광진구 능동로 209, 세종대학교 대양 AI센터 601호(군자동)

한대찬

서울특별시 광진구 광나루로19길 44-1, 201호(군자동)

(74) 대리인

송인호, 윤형근, 최관락

전체 청구항 수 : 총 8 항

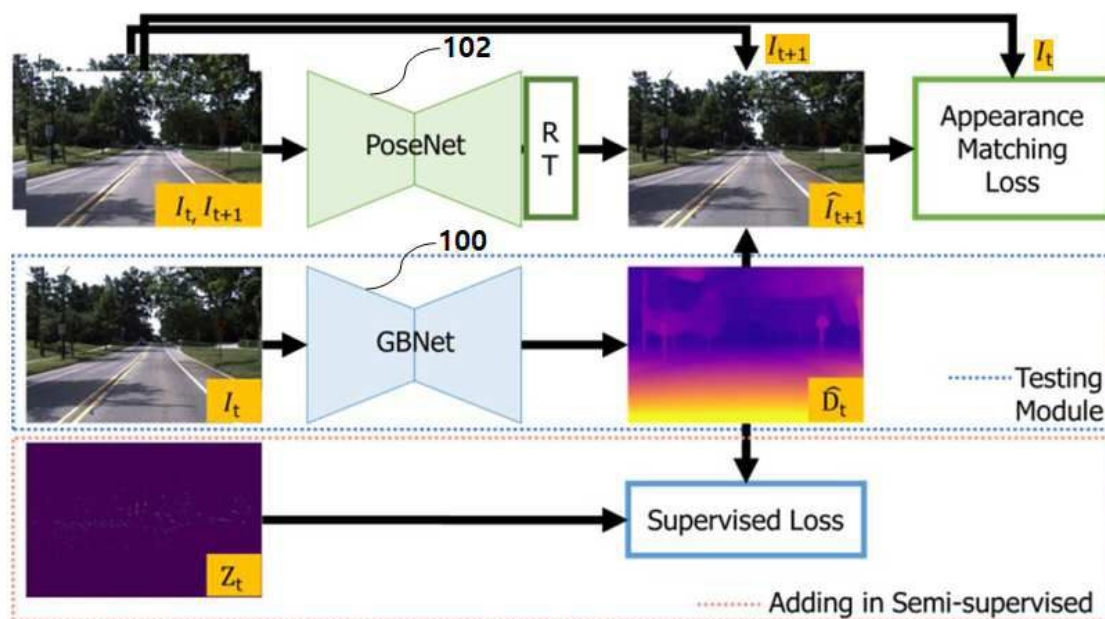
(54) 발명의 명칭 그래디언트 부스팅 네트워크를 이용한 단안 영상 기반 깊이 추정 방법 및 장치

(57) 요약

본 발명은 그래디언트 부스팅 네트워크를 이용한 단안 영상 기반 깊이 추정 방법 및 장치를 개시한다. 본 발명에 따르면, 프로세서 및 상기 프로세서에 연결되는 메모리를 포함하되, 상기 메모리는, 깊이 맵 추정을 위해 순차적으로 연결되는 n개의 단일 네트워크를 포함하는 단안 영상 기반 깊이 추정 모델에 제1 시점의 제1 타겟 영상을

(뒷면에 계속)

대표도 - 도1



입력하여 상기 n개의 단일 네트워크 각각이 상기 제1 타겟 영상으로부터 순차적으로 출력하는 깊이 맵을 통해 최종 깊이 맵을 획득하고, 예고 모션 추정기를 통해 상기 제1 타겟 영상, 상기 제1 시점에 인접한 제2 시점의 제2 타겟 영상 및 상기 최종 깊이 맵을 이용하여 상기 제1 타겟 영상에 대응하는 합성 타겟 영상을 생성하고, 상기 합성 타겟 영상과 상기 제1 타겟 영상 간의 차이를 계산하여 상기 최종 깊이 맵의 추정을 반복적으로 수행하도록, 상기 프로세서에 의해 실행되는 프로그램 명령어들을 저장하는 단안 영상 기반 깊이 추정 장치가 제공된다.

(52) CPC특허분류

G06N 3/04 (2023.01)

G06N 3/08 (2023.01)

G06T 2207/10028 (2013.01)

G06T 2207/20081 (2013.01)

G06T 2207/20084 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711139507
과제번호	2021-0-02067-001
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	차세대인공지능핵심원천기술개발(R&D)
연구과제명	다목적 비디오 검색을 위한 차세대 인공신경망 기술 개발
기 여 율	1/2
과제수행기관명	세종대학교 산학협력단
연구기간	2021.07.01 ~ 2022.03.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711130928
과제번호	2020M3F6A1109603
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	미래국방혁신기술개발(R&D)
연구과제명	전장 환경 인식 및 무기체계 성능을 지원하는 AI 카메라 기술 개발
기 여 율	1/2
과제수행기관명	세종대학교 산학협력단
연구기간	2021.02.01 ~ 2022.01.31

명세서

청구범위

청구항 1

그래디언트 부스팅 네트워크를 이용한 단안 영상 기반 깊이 추정 장치로서,

프로세서; 및

상기 프로세서에 연결되는 메모리를 포함하되,

상기 메모리는,

깊이 맵 추정을 위해 순차적으로 연결되는 n 개의 단일 네트워크를 포함하는 단안 영상 기반 깊이 추정 모델에 제1 시점의 제1 타겟 영상을 입력하여 상기 n 개의 단일 네트워크 각각이 상기 제1 타겟 영상으로부터 순차적으로 출력하는 깊이 맵을 통해 최종 깊이 맵을 획득하고,

예고 모션 추정기를 통해 상기 제1 타겟 영상, 상기 제1 시점에 인접한 제2 시점의 제2 타겟 영상 및 상기 최종 깊이 맵을 이용하여 상기 제1 타겟 영상에 대응하는 합성 타겟 영상을 생성하고,

상기 합성 타겟 영상과 상기 제1 타겟 영상 간의 차이를 계산하여 상기 최종 깊이 맵의 추정을 반복적으로 수행하도록,

상기 프로세서에 의해 실행되는 프로그램 명령어들을 저장하는 단안 영상 기반 깊이 추정 장치.

청구항 2

제1항에 있어서,

상기 n 개의 단일 네트워크는 인코더-디코더 신경망으로 구성되는 PackNet인 단안 영상 기반 깊이 추정 장치.

청구항 3

제1항에 있어서,

상기 프로그램 명령어들은,

상기 n 개의 단일 네트워크 중 첫 번째 단일 네트워크는 상기 제1 타겟 영상을 입력으로 하여 제1 깊이 맵을 추정하고, 나머지 $n-1$ 개의 단일 네트워크 각각은 이전 단일 네트워크에서 추정된 깊이 맵을 입력으로 하여 깊이 맵을 추정하고, 상기 n 개의 단일 네트워크가 출력하는 n 개의 깊이 맵에 서로 다른 가중치를 적용하여 최종 깊이 맵을 추정하는 단안 영상 기반 깊이 추정 장치.

청구항 4

제3항에 있어서,

상기 제1 깊이 맵에 가장 큰 가중치가 부여되고, 상기 나머지 $n-1$ 개의 단일 네트워크 각각이 추정하는 깊이 맵에 대해 순차적으로 작은 가중치가 부여되는 단안 영상 기반 깊이 추정 장치.

청구항 5

제1항에 있어서,

상기 합성 타겟 영상과 상기 제1 타겟 영상 간의 차이는 상기 합성 타겟 영상 및 상기 제1 타겟 영상의 픽셀 수준의 유사도를 높이기 위한 L1 거리 및 구조적 유사성(Structural Similarity, SSIM)의 조합으로 구성되는 외관 일치 손실로 정의되는 단안 영상 기반 깊이 추정 장치.

청구항 6

제1항에 있어서,

상기 최종 깊이 맵에 대해 상기 제1 타겟 영상에 대한 LiDAR 데이터를 이용하여 준 지도 학습이 수행되는 단안 영상 기반 깊이 추정 장치.

청구항 7

프로세서 및 메모리를 포함하는 장치에서 단안 영상 기반으로 깊이를 추정하는 방법으로서,

깊이 맵 추정을 위해 순차적으로 연결되는 n 개의 단일 네트워크를 포함하는 단안 영상 기반 깊이 추정 모델에 제1 시점의 제1 타겟 영상을 입력하여 상기 n 개의 단일 네트워크 각각이 상기 제1 타겟 영상으로부터 순차적으로 출력하는 깊이 맵을 통해 최종 깊이 맵을 획득하는 단계;

예고 모션 추정기를 통해 상기 제1 타겟 영상, 상기 제1 시점에 인접한 제2 시점의 제2 타겟 영상 및 상기 최종 깊이 맵을 이용하여 상기 제1 타겟 영상에 대응하는 합성 타겟 영상을 생성하는 단계; 및

상기 합성 타겟 영상과 상기 제1 타겟 영상 간의 차이를 계산하여 상기 최종 깊이 맵의 추정을 반복적으로 수행하는 단계를 포함하는 단안 영상 기반 깊이 추정 방법.

청구항 8

제7항에 따른 방법을 수행하는 프로그램이 기록된 컴퓨터 판독 가능한 기록매체.

발명의 설명

기술 분야

[0001] 본 발명은 그래디언트 부스팅 네트워크를 이용한 단안 영상 기반 깊이 추정 방법 및 장치에 관한 것이다.

배경 기술

[0002] 단일 영상에서의 깊이 추정(단안 영상 깊이 추정)은 주어진 시점에서 하나의 카메라를 통해 촬영된 2차원 영상으로부터 객체까지의 3차원 거리 정보를 추정하는 것이다.

[0003] 즉, 단안 영상 깊이 추정의 목표는 각 장면 요소가 카메라에서 얼마나 멀리 떨어져 있는지에 대한 픽셀 단위 추정을 위한 깊이 맵을 생성하는 것이다.

[0004] LiDAR) 또는 스테레오 카메라 시스템을 통해 직간접적으로 깊이를 추정할 수 있지만 단안 카메라는 저렴하고 많이 사용되고 있기 때문에 단안 영상 기반 깊이 추정에 대한 다양한 연구가 진행되고 있다.

[0005] 최근 딥러닝 기반으로 단안 RGB 영상에서 깊이 정보 추정에 유용한 특징 맵을 추출하고 이를 이용해서 깊이를 추정하는 모델들이 기존 방법들의 성능을 넘어서면서 관련된 연구가 활발히 진행되고 있다.

[0006] 그러나, 기존 자가 지도 학습 및 준 지도 학습기반 단안 영상 깊이 추정의 경우 단일 딥러닝 네트워크만을 사용해 예측을 하고 있다.

[0007] 따라서 하나의 딥러닝 네트워크에 의존해서 보정되지 못한 깊이 정보만을 이용해야 한다는 문제를 가지고 있다.

선행기술문헌

특허문헌

[0008] (특허문헌 0001) 대한민국공개특허공보 10-2021-0084365

발명의 내용

해결하려는 과제

[0009] 상기한 종래기술의 문제점을 해결하기 위해, 본 발명은 단일 딥러닝 네트워크에서 추정된 깊이 영상을 보완하고 보정하여 더욱 정교한 깊이 추정을 할 수 있는 그래디언트 부스팅 네트워크를 이용한 단안 영상 기반 깊이 추정 방법 및 장치를 제안하고자 한다.

과제의 해결 수단

- [0010] 상기한 바와 같은 목적을 달성하기 위하여, 본 발명의 일 실시예에 따르면, 그래디언트 부스팅 네트워크를 이용한 단안 영상 기반 깊이 추정 장치로서, 프로세서; 및 상기 프로세서에 연결되는 메모리를 포함하되, 상기 메모리는, 깊이 맵 추정을 위해 순차적으로 연결되는 n 개의 단일 네트워크를 포함하는 단안 영상 기반 깊이 추정 모델에 제1 시점의 제1 타겟 영상을 입력하여 상기 n 개의 단일 네트워크 각각이 상기 제1 타겟 영상으로부터 순차적으로 출력하는 깊이 맵을 통해 최종 깊이 맵을 획득하고, 예고 모션 추정기를 통해 상기 제1 타겟 영상, 상기 제1 시점에 인접한 제2 시점의 제2 타겟 영상 및 상기 최종 깊이 맵을 이용하여 상기 제1 타겟 영상에 대응하는 합성 타겟 영상을 생성하고, 상기 합성 타겟 영상과 상기 제1 타겟 영상 간의 차이를 계산하여 상기 최종 깊이 맵의 추정을 반복적으로 수행하도록, 상기 프로세서에 의해 실행되는 프로그램 명령어들을 저장하는 단안 영상 기반 깊이 추정 장치가 제공된다.
- [0011] 상기 n 개의 단일 네트워크는 인코더-디코더 신경망으로 구성되는 PackNet일 수 있다.
- [0012] 상기 프로그램 명령어들은, 상기 n 개의 단일 네트워크 중 첫 번째 단일 네트워크는 상기 제1 타겟 영상을 입력으로 하여 제1 깊이 맵을 추정하고, 나머지 $n-1$ 개의 단일 네트워크 각각은 이전 단일 네트워크에서 추정된 깊이 맵을 입력으로 하여 깊이 맵을 추정하고, 상기 n 개의 단일 네트워크가 출력하는 n 개의 깊이 맵에 서로 다른 가중치를 적용하여 최종 깊이 맵을 추정할 수 있다.
- [0013] 상기 제1 깊이 맵에 가장 큰 가중치가 부여되고, 상기 나머지 $n-1$ 개의 단일 네트워크 각각이 추정하는 깊이 맵에 대해 순차적으로 작은 가중치가 부여될 수 있다.
- [0014] 상기 합성 타겟 영상과 상기 제1 타겟 영상 간의 차이는 상기 합성 타겟 영상 및 상기 제1 타겟 영상의 픽셀 수준의 유사도를 높이기 위한 L1 거리 및 구조적 유사성(Structural Similarity, SSIM)의 조합으로 구성되는 외관 일치 손실로 정의될 수 있다.
- [0015] 상기 최종 깊이 맵에 대해 상기 제1 타겟 영상에 대한 LiDAR 데이터를 이용하여 준 지도 학습이 수행될 수 있다.
- [0016] 본 발명의 다른 측면에 따르면, 프로세서 및 메모리를 포함하는 장치에서 단안 영상 기반으로 깊이를 추정하는 방법으로서, 깊이 맵 추정을 위해 순차적으로 연결되는 n 개의 단일 네트워크를 포함하는 단안 영상 기반 깊이 추정 모델에 제1 시점의 제1 타겟 영상을 입력하여 상기 n 개의 단일 네트워크 각각이 상기 제1 타겟 영상으로부터 순차적으로 출력하는 깊이 맵을 통해 최종 깊이 맵을 획득하는 단계; 예고 모션 추정기를 통해 상기 제1 타겟 영상, 상기 제1 시점에 인접한 제2 시점의 제2 타겟 영상 및 상기 최종 깊이 맵을 이용하여 상기 제1 타겟 영상에 대응하는 합성 타겟 영상을 생성하는 단계; 및 상기 합성 타겟 영상과 상기 제1 타겟 영상 간의 차이를 계산하여 상기 최종 깊이 맵의 추정을 반복적으로 수행하는 단계를 포함하는 단안 영상 기반 깊이 추정 방법이 제공된다.
- [0017] 본 발명의 또 다른 측면에 따르면, 상기한 방법을 수행하는 프로그램이 기록된 컴퓨터 판독 가능한 기록매체가 제공된다.

발명의 효과

- [0018] 본 발명에 따르면, 개별적으로 깊이를 추정하는 n 개의 단일 네트워크를 통과할때마다 점진적으로 깊이 오차를 감소시켜 정확한 깊이 추정이 가능한 장점이 있다.

도면의 간단한 설명

- [0019] 도 1은 본 발명의 바람직한 일 실시예에 따른 단안 영상 기반 깊이 추정을 위한 전체 네트워크를 도시한 도면이다.
- 도 2는 본 실시예에 따른 GBNet의 상세 구성을 도시한 도면이다.
- 도 3은 본 발명의 바람직한 일 실시예에 따른 단안 영상 기반 깊이 추정 장치를 도시한 도면이다.
- 도 4는 DDAD 데이터셋에 대한 자가 및 준 지도 학습을 통한 감독 단안 영상 기반 깊이 추정의 성능을 나타낸 것이다.
- 도 5는 본 실시예에 따른 깊이 추정을 위한 단일 네트워크의 효과를 나타낸 것이다.

발명을 실시하기 위한 구체적인 내용

- [0020] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세하게 설명하고자 한다.
- [0021] 그러나, 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.
- [0023] 본 실시예에서는 단일 딥러닝 네트워크에 예측된 깊이 맵을 점진적으로 보완하여 정확한 깊이 추정이 가능한 아키텍처를 제안한다.
- [0024] 본 실시예에 따른 네트워크는 PackNet을 기반으로 하며, 그래디언트 부스팅(Gradient Boosting) 기법을 적용하여 단일 네트워크인 PackNet를 반복적으로 사용하여 각 PackNet의 출력을 보정하여 노이즈를 제거하고 더욱 선명하게 깊이 맵을 추정한다.
- [0025] 도 1은 본 발명의 바람직한 일 실시예에 따른 단안 영상 기반 깊이 추정을 위한 전체 네트워크를 도시한 도면이다.
- [0026] 도 1에 도시된 바와 같이, 본 실시예에 따른 단안 영상 기반 깊이 추정을 위한 네트워크는 타겟 영상(I_t)으로부터 깊이를 추정하는 단안 깊이 추정 모델(100) 및 I_t 의 포즈에 대해, 각 소스 영상 $I_{t+1} \in [I_t; I_{t+1}]$ 의 개별 포즈 $e_{t \rightarrow t+1}$ 를 추정하는 에고 모션(ego motion) 추정기(PoseNet, 102)를 포함할 수 있다.
- [0027] 단안 깊이 추정 모델(100)은 순차적으로 연결되는 n 개의 단일 네트워크를 포함하며, 제1 시점의 제1 타겟 영상(I_t)으로부터 n 개의 단일 네트워크가 순차적으로 출력하는 깊이 맵을 통해 최종 깊이 맵 $\hat{D}_t$ 를 출력한다.
- [0028] 본 실시예에 따른 단안 영상 기반 깊이 추정은 조밀한 깊이 맵을 사용하지 않는 자가/준(self-/semi-) 지도 학습 방법을 이용한다.
- [0029] 본 실시예에 따른 단안 깊이 추정 모델(100)은 깊이 추정을 위해 순차적으로 연결되는 n 개의 단일 네트워크를 포함하는 깊이 맵의 오차를 순차적으로 감소시킨다는 점에서 그래디언트 부스팅 네트워크(GBNet)으로 정의된다.
- [0030] 도 2는 본 실시예에 따른 GBNet의 상세 구성을 도시한 도면이다.
- [0031] 도 2를 참조하면, 본 실시예에 따른 GBNet은 n 개의 단일 네트워크 P_i 를 통과할때마다 점진적으로 깊이 오차가 감소하는 구조를 가진다.
- [0032] 본 실시예에 따른 단일 네트워크는 PackNet일 수 있다.
- [0033] PackNet은 새로운 패킹(packing) 및 언패킹 블록을 활용하여 중요한 공간 세부 정보를 보존하는 방법을 학습하여 고품질 깊이 예측으로 이어지는 인코더-디코더 신경망이다.
- [0034] 도 2를 참조하면, 첫 번째 제1 단일 네트워크 $P_1(200-1)$ 는 제1 시점(t)의 타겟 영상 I_t 를 입력으로 하고, 나머지 $n-1$ 개의 단일 네트워크 $P_i(200-2$ 내지 $200-n)$ 는 이전 단일 네트워크의 출력 $\hat{D}_{t(i-1)}$ 를 입력으로 한다.
- [0035] 모든 영상의 풍부한 정보를 활용하고 상호 보완적인 관계를 만들기 위해 최종 깊이 맵 $\hat{D}_t$ 를 다음과 같이 추정한다.

수학식 1

$$\hat{D}_t = \sum_{i=1}^N \lambda_i \hat{D}_{ti},$$

[0036]

[0037] 여기서, λ 는 각 단일 네트워크의 영향을 조정하여 보다 정교한 깊이 맵의 효과를 증가시키기 위한 가중치이며, i 가 작을수록 큰 값을 가질 수 있다.

[0038] 이하에서는 본 실시예에 따른 단안 영상 깊이 추정 네트워크의 자가 지도 학습 과정을 설명한다.

[0039] 상기한 바와 같이, GBNet(100)을 통해 단안 영상 기반 깊이 추정이 얻어진 후 깊이 및 포즈 모델(에고 모션 추정기)의 동시 학습이 수행된다.

[0040] 본 실시예에 따르면, 제1 시점(t)의 제1 타겟 영상, 제1 시점에 인접한 제2 시점($t+1$)의 제2 타겟 영상 및 상기 한 최종 깊이 맵을 이용하여 제1 타겟 영상에 대응되는 합성 타겟 영상($\hat{I}_{t+1}$)을 생성하고, 합성 타겟 영상과 상기 제1 타겟 영상의 차이가 최소가 되도록 최종 깊이 맵의 추정을 반복적으로 수행된다.

[0041] 이와 같은 과정이 자가 지도 학습으로 정의될 수 있다.

[0042] 합성 타겟 영상과 제1 타겟 영상 간의 차이는 합성 타겟 영상 및 제1 타겟 영상의 픽셀 수준의 유사도를 높이기 위한 L1 거리 및 구조적 유사성(Structural Similarity, SSIM)의 조합으로 구성되는 외관 일치 손실(appearance matching loss)로 정의될 수 있다.

수학식 2

$$L_{sl}^p(I_t, \hat{I}_{t+1}) = \alpha \frac{1 - SSIM(I_t, \hat{I}_{t+1})}{2} + (1 - \alpha) \|I_t - \hat{I}_{t+1}\|$$

[0043]

[0044] 수학식 2는 일반적으로 자가 지도를 위한 강건한 학습 방법이다.

[0045] 본 실시예에 따른 자가 지도 학습 제약 조건은 합성 타겟 영상 $\hat{I}_{t+1}$ 과 타겟 영상 I_t 및 깊이 평활도 손실(Depth Smoothness Loss) 항 L_{sl}^s 가 추정된 깊이 $\hat{D}_t$ 를 국소적으로 부드러워지도록 하기 위한 외관 일치 손실(appearance matching loss) 항 L_{sl}^p 로 구성된다.

[0046] 그러나 장면에서의 시차 오차로 인해 시야를 벗어나고 가려진 픽셀이 생성되며, 이는 학습에 바람직하지 않은 영향을 미친다.

[0047] 픽셀당 최소 재투사 손실을 이용하여 시야를 벗어난 픽셀 및 가려진 픽셀 문제를 완화할 수 있다. 이는 각 소스 영상 I_{t+1} 에 대한 픽셀당 최소 손실을 계산하여 바람직하지 않은 문제를 완화한다.

[0048] 이는 동일한 픽셀이 시야를 벗어나 합성 타겟 영상 $\hat{I}_{t+1}$ 에 가려지지 않음을 의미한다.

수학식 3

[0049]
$$L_{sl}^p(I_t, I_{t+1}) = \min L_{sl}^p(I_t, \hat{I}_{t+1})$$

[0050] 또한, 본 실시예에서는 자동 마스킹 정적 픽셀 기법(auto-masking static pixels method)을 적용한다.

[0051] 정적 픽셀로 인해 약간의 일치 손실이 있고 프레임 사이에 예고 모션이 없다고 가정할 때 무한한 깊이 홀을 만들 수 있으므로 자동 마스킹을 사용하여 정적 픽셀을 무시한다.

[0052] 본 실시예에서는 마스크를 생성하기 위해 $L_{sl}^p(I_t, \hat{I}_{t+1})$ 보다 더 높은 $L_{sl}^p(I_t, I_{t+1})$ 를 갖는 픽셀을 탐색한다.

수학식 4

[0053]
$$M = \min L_{sl}^p(I_t, I_{t+1}) > \min L_{sl}^p(I_t, \hat{I}_{t+1})$$

[0054] 깊이 평활도 손실(Depth Smoothness Loss)은 텍스처가 없는 낮은 영상 그래디언트 영역에서 깊이 불연속성에 패널티를 부여하고, 제약 조건에 깊이 평활도를 적용한다.

수학식 5

[0055]
$$L_{sl}^s = |\delta_x \hat{D}_t| e^{-|\delta_x I_t|} + |\delta_y \hat{D}_t| e^{-|\delta_y I_t|}$$

[0056] 프로세스 전반에 걸친 자가 지도 손실은 다음과 같다.

수학식 6

[0057]
$$L_{sl} = L_{sl}^p(I_t, I_{t+1}) \odot M + 0.001 * L_{sl}^s$$

[0058] 여기서, $\odot$ 는 요소별 곱을 나타낸다.

[0059] 도 1에 도시된 바와 같이, 본 실시예에서는 네트워크에 정확한 깊이 정보를 제공하기 위해 LiDAR 데이터 Z를 이용한 준 지도 학습을 사용한다.

[0060] 자가 지도 목표보다 더 자세한 추정을 유도하기 위해 회소 깊이 정보를 포함하는 LiDAR 데이터 Z는 지도 손실의 정답(ground truth)이다. 지도 학습의 제약 조건은 픽셀에서 사용 가능한 정답으로부터 추론된 깊이 맵의 편차를 측정한다.

수학식 7

$$L_{sp}(I_t, Z_t) = \|G(I_t) - z_t\|, z_t = Z_t > 0$$

지도 (L_{sp}) 및 준 지도 (L_{sl}) 목표를 포함하는 전체 준 지도 손실 (L_{sm}) 함수는 다음과 같다.

수학식 8

$$L_{sm}(I_t, I_{t+1}, Z_t, Z_{t+1}) = \gamma L_{sp}(I_t, Z_t) + \beta L_{sl}(I_t, I_{t+1})$$

where, $\gamma = 0.9$ and $\beta = 0.1$.

도 3은 본 발명의 바람직한 일 실시예에 따른 단안 영상 기반 깊이 추정 장치를 도시한 도면이다.

도 3에 도시된 바와 같이, 본 실시예에 따른 장치는 프로세서(300) 및 메모리(302)를 포함할 수 있다.

프로세서(300)는 컴퓨터 프로그램을 실행할 수 있는 CPU(central processing unit)나 그 밖에 가상 머신 등을 포함할 수 있다.

메모리(302)는 고정식 하드 드라이브나 착탈식 저장 장치와 같은 불휘발성 저장 장치를 포함할 수 있다. 착탈식 저장 장치는 콤팩트 플래시 유닛, USB 메모리 스틱 등을 포함할 수 있다. 메모리(302)는 각종 랜덤 액세스 메모리와 같은 휘발성 메모리도 포함할 수 있다.

본 발명의 일 실시예에 따르면, 메모리(302)에는 복수의 단일 네트워크를 이용하여 단안 영상 기반 깊이 추정을 위한 프로그램 명령어들이 저장된다.

본 실시예에 따른 프로그램 명령어들은, 깊이 맵 추정을 위해 순차적으로 연결되는 n개의 단일 네트워크를 포함하는 단안 영상 기반 깊이 추정 모델에 제1 시점의 제1 타겟 영상을 입력하여 상기 n개의 단일 네트워크 각각이 상기 제1 타겟 영상으로부터 순차적으로 출력하는 깊이 맵을 통해 최종 깊이 맵을 획득하고, 예고 모션 추정기를 통해 상기 제1 타겟 영상, 상기 제1 시점에 인접한 제2 시점의 제2 타겟 영상 및 상기 최종 깊이 맵을 이용하여 상기 제1 타겟 영상에 대응하는 합성 타겟 영상을 생성하고, 상기 합성 타겟 영상과 상기 제1 타겟 영상 간의 차이를 계산하여 상기 최종 깊이 맵의 추정을 반복적으로 수행한다.

여기서, n개의 단일 네트워크의 깊이 맵 추정 과정 및 가중치에 설정에 대한 내용은 이미 상술하였으므로 이에 대한 상세한 설명은 생략한다.

도 4는 DDAD 데이터셋에 대한 자가 및 준 지도 학습을 통한 감독 단안 영상 기반 깊이 추정의 성능을 나타낸 것이다.

도 4에 나타난 바와 같이, 기존 PackNet에 비해 본 실시예에 따른 GBNet을 이용하는 경우 깊이 추정의 정확도가 높아지는 것을 확인할 수 있다.

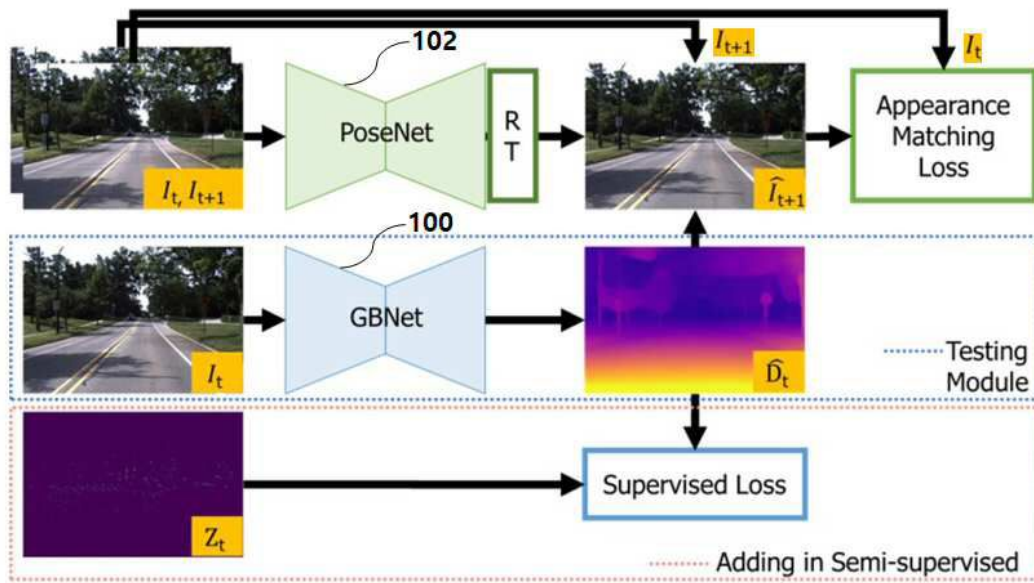
도 5는 본 실시예에 따른 깊이 추정을 위한 단일 네트워크의 효과를 나타낸 것이다.

도 5는 깊이 추정을 위한 3개의 단일 네트워크를 순차적으로 연결한 경우를 나타낸 것으로, 여러 개의 단일 네트워크를 순차적으로 연결하여 깊이 추정의 정확도가 높아지는 것을 확인할 수 있다.

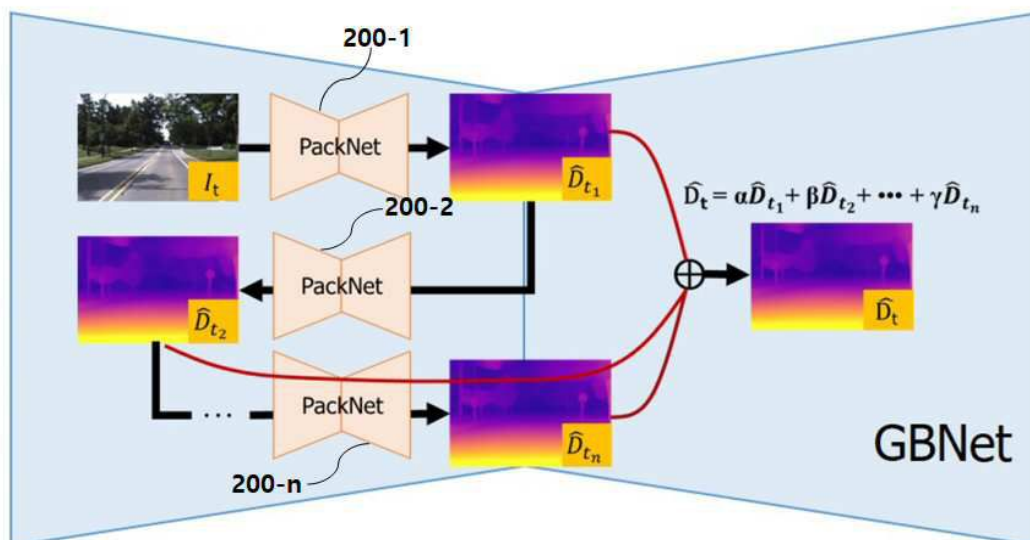
상기한 본 발명의 실시예는 예시의 목적을 위해 개시된 것이고, 본 발명에 대한 통상의 지식을 가지는 당업자라면 본 발명의 사상과 범위 안에서 다양한 수정, 변경, 부가가 가능할 것이며, 이러한 수정, 변경 및 부가는 하기의 특허청구범위에 속하는 것으로 보아야 할 것이다.

도면

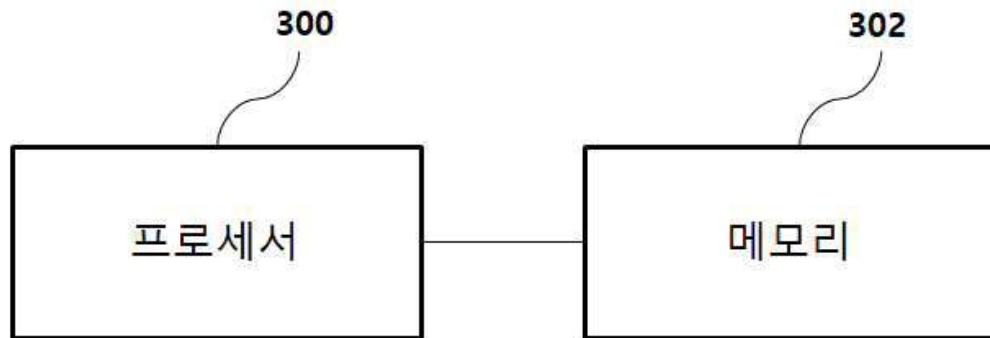
도면1



도면2



도면3



도면4

Type	Method	Abs Rel ↓	Sqr Rel ↓	RMSE ↓	RMSE log ↓	$\delta_{1.25} \uparrow$	$\delta_{1.25^2} \uparrow$	$\delta_{1.25^3} \uparrow$
self-supervised	PackNet	0.175	5.002	15.326	0.263	0.793	0.921	0.962
	GBNet_{total}	0.148	3.329	14.471	0.244	0.818	0.930	0.967
semi-supervised	PackNet	0.122	2.477	13.200	0.220	0.849	0.941	0.973
	GBNet_{total}	0.124	2.476	13.276	0.220	0.846	0.940	0.973

도면5

Type	Method	Abs Rel ↓	Sqr Rel ↓	RMSE ↓	RMSE log ↓	$\delta_{1.25} \uparrow$	$\delta_{1.25^2} \uparrow$	$\delta_{1.25^3} \uparrow$
self-supervised	GBNet ₁	0.254	6.055	17.582	0.331	0.606	0.871	0.950
	GBNet ₂	0.150	3.394	14.661	0.246	0.812	0.928	0.967
	GBNet ₃	0.164	3.467	15.148	0.256	0.786	0.923	0.966
	GBNet_{total}	0.148	3.329	14.471	0.244	0.818	0.930	0.967
semi-supervised	GBNet ₁	0.191	4.411	16.757	0.275	0.740	0.913	0.962
	GBNet ₂	0.130	2.609	13.335	0.223	0.841	0.939	0.973
	GBNet ₃	0.143	2.855	14.460	0.238	0.817	0.929	0.969
	GBNet_{total}	0.124	2.476	13.276	0.220	0.846	0.940	0.973