(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0097629
(43) 공개일자 2022년07월08일

(51) 국제특허분류(Int. Cl.)
*G06T 7/55* (2017.01) *G06N 20/00* (2019.01)
*G06T 7/00* (2017.01)
(52) CPC특허분류
*G06T 7/55* (2017.01)
*G06N 20/00* (2021.08)
(21) 출원번호 10-2020-0187626
(22) 출원일자 2020년12월30일
심사청구일자 2020년12월30일

(71) 출원인
**한국기술교육대학교 산학협력단**
충청남도 천안시 동남구 병천면 충절로 1600 (한국기술교육대학교내)
(72) 발명자
**주영복**
충청남도 천안시 서북구 충무로 124-24, 116동 904호(쌍용동 현대아이파크)
**박혜승**
충청남도 천안시 서북구 공원로 195, 102동 2404호 (불당동, 펜타포트)
(74) 대리인
**특허법인(유한) 대아**

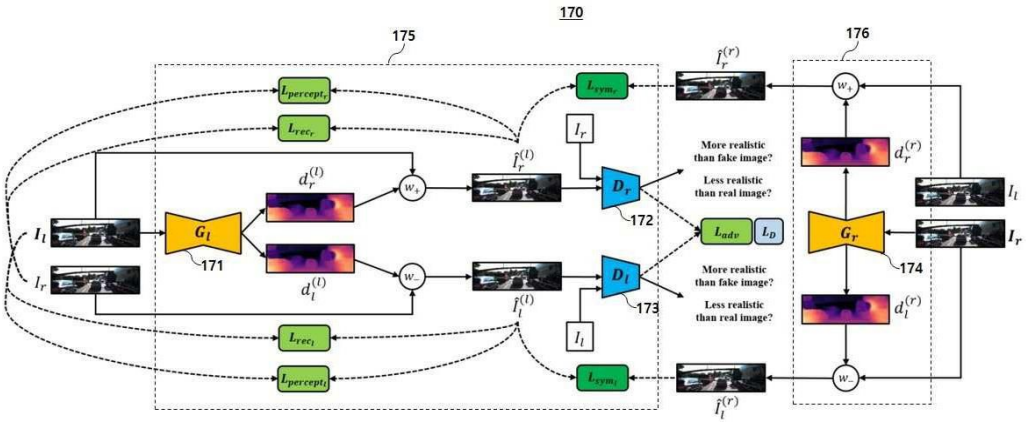
전체 청구항 수 : 총 5 항

(54) 발명의 명칭 **비지도 학습 기반의 깊이 예측 방법 및 이를 이용하는 시스템**

## (57) 요 약

본 발명의 일 실시 예에 따른 비지도 학습 기반의 깊이 예측 시스템은, 실제 이미지를 이용하여 페이크 이미지를 만드는 생성자; 및 실제 이미지와 비교하여 상기 페이크 이미지를 판별하는 판별자(discriminator)를 포함하고, 판별자는, 훈련 손실 함수를 이용하여, 생성자에 의해 만들어진 페이크 이미지에 대해, 페이크 이미지가 실제 이미지보다 얼마나 덜 현실적인지, 그리고 실제 이미지가 페이크 이미지보다 얼마나 더 현실적인지에 대한 확률을 출력한다. 본 발명에 의하면, 또한, 합리적인 훈련 오버헤드와 우측 이미지에 대한 간접적인 학습을 위한 대칭적 손실을 통해 예측되는 시차맵의 정확성이 향상될 수 있다.

## 대 표 도



(52) CPC특허분류
***G06T 7/001*** (2013.01)
*G06T 2207/10028* (2013.01)
*G06T 2207/20076* (2013.01)
*G06T 2207/20081* (2013.01)

이 발명을 지원한 국가연구개발사업

| | |
|---|---|
| 과제고유번호 | 1345323366 |
| 과제번호 | LINCPLUS-2020-45 |
| 부처명 | 교육부 |
| 과제관리(전문)기관명 | 한국연구재단 |
| 연구사업명 | 사회맞춤형산학협력선도대학육성(LINC+) |
| 연구과제명 | 단안 영상의 깊이 정보 추정을 위한 준지도 학습 기반 딥러닝 프레임워크 개발 |
| 기 여 율 | 1/1 |
| 과제수행기관명 | 한국기술교육대학교산학협력단 |
| 연구기간 | 2020.06.01 ~ 2020.12.31 |

# 명 세 서

## 청구범위

### 청구항 1

실제 이미지를 이용하여 페이크 이미지를 만드는 생성자(generator); 및

상기 실제 이미지와 비교하여 상기 페이크 이미지를 판별하는 판별자(discriminator)를 포함하고,

상기 판별자는,

훈련 손실 함수를 이용하여, 상기 생성자에 의해 만들어진 페이크 이미지에 대해, 상기 페이크 이미지가 상기 실제 이미지보다 얼마나 덜 현실적인지, 그리고 상기 실제 이미지가 상기 페이크 이미지보다 얼마나 더 현실적인지에 대한 확률을 출력하는 것을 특징으로 하는,

비지도 학습 기반의 깊이 예측 시스템.

### 청구항 2

좌측 실제 이미지 및 우측 실제 이미지 중에서 어느 하나를 이용하여 좌측 페이크 이미지 및 우측 페이크 이미지를 만드는 생성자(generator); 및

상기 좌측 실제 이미지 또는 우측 실제 이미지와 비교하여 상기 좌측 페이크 이미지와 우측 페이크 이미지를 각각 판별하는 2 개의 상대론적 판별자(relativistic discriminator)를 포함하고,

상기 판별자는,

상기 생성자에 의해 만들어진 페이크 이미지에 대해, 상기 페이크 이미지가 상기 실제 이미지보다 얼마나 덜 현실적인지, 그리고 상기 실제 이미지가 상기 페이크 이미지보다 얼마나 더 현실적인지에 대한 확률을 출력하는 것을 특징으로 하는,

비지도 학습 기반의 깊이 예측 시스템.

### 청구항 3

메인 네트워크 및 서브 네트워크를 포함하고,

상기 메인 네트워크는,

좌측 실제 이미지 및 우측 실제 이미지 중에서 어느 하나를 이용하여 제1 쌍의 좌측 페이크 이미지 및 우측 페이크 이미지를 만드는 제1 생성자(generator); 및

상기 좌측 실제 이미지 또는 우측 실제 이미지와 비교하여 상기 좌측 페이크 이미지와 우측 페이크 이미지를 각각 판별하는 2 개의 상대론적 판별자(relativistic discriminator)를 포함하고,

상기 서브 네트워크는,

상기 좌측 이미지 및 우측 이미지 중에서 다른 하나를 이용하여 제2 쌍의 좌측 페이크 이미지 및 우측 페이크 이미지를 생성하는 제2 생성자를 포함하되,

상기 판별자는,

상기 제1 쌍의 페이크 이미지 및 제2 쌍의 페이크 이미지 간의 대칭적 손실을 포함하는 훈련 손실 함수를 이용하여, 상기 제1 생성자에 의해 만들어진 페이크 이미지에 대해, 상기 페이크 이미지가 상기 실제 이미지보다 얼마나 덜 현실적인지, 그리고 상기 실제 이미지가 상기 페이크 이미지보다 얼마나 더 현실적인지에 대한 확률을 출력하는 것을 특징으로 하는,

비지도 학습 기반의 깊이 예측 시스템.

### 청구항 4

메인 네트워크 및 서브 네트워크를 포함하고,

상기 메인 네트워크는,

좌측 실제 이미지 및 우측 실제 이미지 중에서 어느 하나를 이용하여 제1 쌍의 좌측 페이크 이미지 및 우측 페이크 이미지를 만드는 제1 생성자(generator); 및

상기 좌측 실제 이미지 또는 우측 실제 이미지와 비교하여 상기 좌측 페이크 이미지와 우측 페이크 이미지를 각각 판별하는 2 개의 상대론적 판별자(relativistic discriminator)를 포함하고,

상기 서브 네트워크는,

상기 좌측 이미지 및 우측 이미지 중에서 다른 하나를 이용하여 제2 쌍의 좌측 페이크 이미지 및 우측 페이크 이미지를 생성하는 제2 생성자를 포함하고,

상기 판별자는,

훈련 손실 함수를 이용하여, 상기 페이크 이미지가 실제 이미지보다 얼마나 덜 현실적인지, 그리고 상기 실제 이미지가 상기 페이크 이미지보다 얼마나 더 현실적인지에 대한 확률을 출력하되,

상기 훈련 손실 함수는,

적대적 손실(Adversarial Loss), GMSD(Gradient Magnitude similarity)를 이용한 이미지 재구성 손실(Image Reconstruction Loss), 인지적 손실(Perceptual Loss), 상기 제1 쌍의 페이크 이미지 및 제2 쌍의 페이크 이미지 간의 대칭적 손실(Symmetric Loss), 좌우 일관성(Left-right Consistency Loss) 손실, 및 Edge Aware Smoothness Loss 중에서 적어도 하나 이상을 포함하도록 구성되는 것을 특징으로 하는,

비지도 학습 기반의 깊이 예측 시스템.

**청구항 5**

실제 이미지인 좌측 이미지 및 우측 이미지 중 어느 하나에 기반하여 제1 쌍의 페이크 이미지를 생성하는 단계;

상기 실제 이미지 중에서 다른 하나에 기반하여 제2 쌍의 페이크 이미지를 생성하는 단계; 및

훈련 손실 함수를 이용하여, 상기 제1 또는 제2 쌍의 페이크 이미지가 상기 실제 이미지보다 얼마나 덜 현실적인지, 그리고 상기 실제 이미지가 상기 제1 또는 제2 쌍의 페이크 이미지보다 얼마나 더 현실적인지에 대한 확률을 출력하는 단계를 포함하되,

상기 훈련 손실 함수는, 좌측 이미지와 우측 이미지 간의 대칭 손실을 포함하도록 구성되는,

비지도 학습 기반의 깊이 예측 방법.

## *발명의 설명*

### *기 술 분 야*

[0001] 본 발명은 비지도 학습 기반의 깊이 예측 방법 및 이를 이용하는 시스템에 관한 것으로, 더욱 상세하게는 단안 이미지를 이용하여 영상의 깊이를 예측하는 방법 및 이를 이용하는 생성적 적대 신경망(Generative Adversarial Network, GAN)을 포함하는 시스템에 관한 것이다.

### *배 경 기 술*

[0002] 싱글 이미지로부터 깊이 예측(depth prediction)은 컴퓨터 비전 영역에서 깊은 역사를 갖는 과제이다. 아무런 힌트 없이 싱글 2D 이미지에서 3D 데이터를 추출하는 것은 어려운 과제이다. 그러나 우리가 정확하게 예측할 수 있다면, 이 추출은 컴퓨터 그래픽에서 합성체(synthetic objects) 삽입, 컴퓨터 사진술에서 합성 깊이, 로봇-보조 수술 및 자동 2D~3D 영화 변환과 같은 다양한 분야에 사용될 수 있다. 하나 이상의 카메라로부터 정확한 깊이 데이터 또한 스마트 감시 시스템에 필수적이고, 값비싼 레이저 기반 시스템을 갖춘 자율주행 자동차도 필수적이다. 그래서 정복해야 할 다양한 연구들이 있어 왔다.

[0003] 관련 기술로서 공개 특허공보 제10-2020-0060194호는 차선들의 깊이 값을 예측하는 방법, 3차원 차선들을 출력

하는 방법 및 그 장치를 개시하고 있다. 그런데 관련 기술에 따른 3차원 차선들을 출력하는 방법은, 제1 영상 및 제2 영상을 모두 이용하여 차선 확률맵을 생성하는 점에서, 단안 영상을 입력 받아 깊이 예측의 학습 모델을 이용하여 영상의 깊이를 예측하는 본 발명의 구성과 구별된다.

## 선행기술문헌

### 특허문헌

[0004] (특허문헌 0001) KR 제10-2020-0060194호 A1 (2020.05.29. 공개)

## 발명의 내용

### 해결하려는 과제

[0005] 본 발명이 해결하고자 하는 일 과제는, 적대적 깊이 예측 네트워크를 이용하여 영상의 깊이를 예측하는 방법 및 시스템을 제공하는 것이다.

[0006] 본 발명이 해결하고자 하는 일 과제는, 적대적 깊이 예측 모델에 대한 각종 특징적인 훈련 손실 함수를 이용하여 영상의 깊이를 예측하는 방법 및 시스템을 제공하는 것이다.

[0007] 본 발명이 해결하고자 하는 일 과제는, 상대론적 판별자가 포함된 생성적 적대적 네트워크를 이용하여 영상의 깊이를 예측하는 방법 및 시스템을 제공하는 것이다.

[0009] 다만, 본 발명이 해결하고자 하는 과제는 이상에서 언급한 것으로 제한되지 않으며, 언급되지 않은 또 다른 해결하고자 하는 과제는 아래의 기재로부터 본 발명이 속하는 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

### 과제의 해결 수단

[0010] 상기와 같은 과제를 해결하기 위한 본 발명의 일 실시 예에 따른 비지도 학습 기반의 깊이 예측 시스템은, 실제 이미지를 이용하여 페이크 이미지를 만드는 생성자(generator); 및 실제 이미지와 비교하여 상기 페이크 이미지를 판별하는 판별자(discriminator)를 포함하고, 판별자는, 훈련 손실 함수를 이용하여, 상기 생성자에 의해 만들어진 페이크 이미지에 대해, 상기 페이크 이미지가 상기 실제 이미지보다 얼마나 덜 현실적인지, 그리고 상기 실제 이미지가 상기 페이크 이미지보다 얼마나 더 현실적인지에 대한 확률을 출력하는 것을 특징으로 한다.

[0011] 본 발명의 일 실시 예에 따른 비지도 학습 기반의 깊이 예측 시스템은, 좌측 실제 이미지 및 우측 실제 이미지 중에서 어느 하나를 이용하여 좌측 페이크 이미지 및 우측 페이크 이미지를 만드는 생성자(generator); 및 좌측 실제 이미지 또는 우측 실제 이미지와 비교하여 상기 좌측 페이크 이미지와 우측 페이크 이미지를 각각 판별하는 2 개의 상대론적 판별자(relativistic discriminator)를 포함하고, 판별자는, 생성자에 의해 만들어진 페이크 이미지에 대해, 상기 페이크 이미지가 상기 실제 이미지보다 얼마나 덜 현실적인지, 그리고 상기 실제 이미지가 상기 페이크 이미지보다 얼마나 더 현실적인지에 대한 확률을 출력하는 것을 특징으로 한다.

[0012] 본 발명의 일 실시 예에 따른 비지도 학습 기반의 깊이 예측 시스템은 , 메인 네트워크 및 서브 네트워크를 포함하고, 메인 네트워크는, 좌측 실제 이미지 및 우측 실제 이미지 중에서 어느 하나를 이용하여 제1 쌍의 좌측 페이크 이미지 및 우측 페이크 이미지를 만드는 제1 생성자(generator); 및 좌측 실제 이미지 또는 우측 실제 이미지와 비교하여 상기 좌측 페이크 이미지와 우측 페이크 이미지를 각각 판별하는 2 개의 상대론적 판별자(relativistic discriminator)를 포함하고, 서브 네트워크는, 좌측 이미지 및 우측 이미지 중에서 다른 하나를 이용하여 제2 쌍의 좌측 페이크 이미지 및 우측 페이크 이미지를 생성하는 제2 생성자를 포함하되, 판별자는, 제1 쌍의 페이크 이미지 및 제2 쌍의 페이크 이미지 간의 대칭적 손실을 포함하는 훈련 손실 함수를 이용하여, 상기 제1 생성자에 의해 만들어진 페이크 이미지에 대해, 상기 페이크 이미지가 상기 실제 이미지보다 얼마나 덜 현실적인지, 그리고 상기 실제 이미지가 상기 페이크 이미지보다 얼마나 더 현실적인지에 대한 확률을 출력하는 것을 특징으로 한다.

[0013] 본 발명의 일 실시 예에 따른 비지도 학습 기반의 깊이 예측 시스템은, 메인 네트워크 및 서브 네트워크를 포함

하고, 메인 네트워크는, 좌측 실제 이미지 및 우측 실제 이미지 중에서 어느 하나를 이용하여 제1 쌍의 좌측 페이크 이미지 및 우측 페이크 이미지를 만드는 제1 생성자(generator); 및 좌측 실제 이미지 또는 우측 실제 이미지와 비교하여 상기 좌측 페이크 이미지와 우측 페이크 이미지를 각각 판별하는 2 개의 상대론적 판별자(relativistic discriminator)를 포함하고, 서브 네트워크는, 좌측 이미지 및 우측 이미지 중에서 다른 하나를 이용하여 제2 쌍의 좌측 페이크 이미지 및 우측 페이크 이미지를 생성하는 제2 생성자를 포함하고, 판별자는, 훈련 손실 함수를 이용하여, 상기 페이크 이미지가 실제 이미지보다 얼마나 덜 현실적인지, 그리고 상기 실제 이미지가 상기 페이크 이미지보다 얼마나 더 현실적인지에 대한 확률을 출력하되, 훈련 손실 함수는, 적대적 손실(Adversarial Loss), GMSD(Gradient Magnitude similarity)를 이용한 이미지 재구성 손실(Image Reconstruction Loss), 인지적 손실(Perceptual Loss), 상기 제1 쌍의 페이크 이미지 및 제2 쌍의 페이크 이미지 간의 대칭적 손실(Symmetric Loss), 좌우 일관성(Left-right Consistency Loss) 손실, 및 Edge Aware Smoothness Loss 중에서 적어도 하나 이상을 포함하도록 구성되는 것을 특징으로 한다.

[0014] 본 발명의 일 실시 예에 따른 비지도 학습 기반의 깊이 예측 방법은, 실제 이미지인 좌측 이미지 및 우측 이미지 중 어느 하나에 기반하여 제1 쌍의 페이크 이미지를 생성하는 단계; 실제 이미지 중에서 다른 하나에 기반하여 제2 쌍의 페이크 이미지를 생성하는 단계; 및 훈련 손실 함수를 이용하여, 제1 또는 제2 쌍의 페이크 이미지가 상기 실제 이미지보다 얼마나 덜 현실적인지, 그리고 실제 이미지가 상기 제1 또는 제2 쌍의 페이크 이미지보다 얼마나 더 현실적인지에 대한 확률을 출력하는 단계를 포함하되, 훈련 손실 함수는, 좌측 이미지와 우측 이미지 간의 대칭 손실을 포함하도록 구성될 수 있다.

[0016] 기타 실시예의 구체적인 사항은 "발명을 실시하기 위한 구체적인 내용" 및 첨부 "도면"에 포함되어 있다.

[0017] 본 발명의 이점 및/또는 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 상세하게 후술되어 있는 각종 실시 예를 참조하면 명확해질 것이다.

[0018] 그러나, 본 발명은 이하에서 개시되는 각 실시예의 구성만으로 한정되는 것이 아니라 서로 다른 다양한 형태로도 구현될 수도 있으며, 단지 본 명세서에서 개시한 각각의 실시 예는 본 발명의 개시가 완전하도록 하며, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에게 본 발명의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 발명은 청구범위의 각 청구항의 범주에 의해 정의될 뿐임을 알아야 한다.

## *발명의 효과*

[0019] 본 발명에 의하면, 향상된 생성적 적대적 네트워크를 이용하여 단안 영상으로부터 고품질의 양안의 스테레오 이미지를 획득할 수 있다.

[0020] 또한, 각종 특징적인 생성적 적대적 네트워크의 손실 함수를 이용하여 실제 영상에 근접한 품질의 가짜 이미지가 생성될 수 있다.

[0021] 또한, 합리적인 훈련 오버헤드와 우측 이미지에 대한 간접적인 학습을 위한 대칭적 손실을 통해 예측되는 시차맵의 정확성이 향상될 수 있다.

## *도면의 간단한 설명*

[0022] 도 1은 본 발명의 일 실시 예에 따른 비지도 학습 기반의 깊이 측정 시스템의 네트워크 예시도이다.

도 2는 본 발명의 일 실시 예에 따른 비지도 학습 기반의 깊이 측정 시스템의 블록도이다.

도 3은 본 발명의 일 실시 예에 따른 비지도 학습 기반의 깊이 측정 방법의 흐름도이다.

도 4는 본 발명의 일 실시 예에 따른 비지도 학습 기반의 깊이 측정 방법의 흐름도이다.

도 5는 본 발명의 일 실시 예에 따른 단안 영상의 깊이 예측을 위해 제안된 비지도 학습 기반의 GAN 신경망의 예시도이다.

도 6은 본 발명의 일 실시 예에 따른 예측된 시차맵을 이용한 이미지 재구성의 예시도이다.

도 7은 본 발명의 일 실시 예에 따른 판별자 네트워크의 블록도이다.

## *발명을 실시하기 위한 구체적인 내용*

[0023] 이하, 첨부한 도면을 참고로 하여 본 발명의 바람직한 실시 예에 대하여 상세히 설명하면 다음과 같다.

[0024] 본 발명을 상세하게 설명하기 전에, 본 명세서에서 사용된 용어나 단어는 통상적이거나 사전적인 의미로 무조건 한정하여 해석되어서는 아니되며, 본 발명의 발명자가 자신의 발명을 가장 최선의 방법으로 설명하기 위해서 각종 용어의 개념을 적절하게 정의하여 사용할 수 있고, 더 나아가 이들 용어나 단어는 본 발명의 기술적 사상에 부합하는 의미와 개념으로 해석되어야 함을 알아야 한다.

[0025] 즉, 본 명세서에서 사용된 용어는 본 발명의 바람직한 실시 예를 설명하기 위해서 사용되는 것일 뿐이고, 본 발명의 내용을 구체적으로 한정하려는 의도로 사용된 것이 아니며, 이들 용어는 본 발명의 여러 가지 가능성을 고려하여 정의된 용어임을 알아야 한다.

[0026] 또한, 본 명세서에 있어서, 단수의 표현은 문맥상 명확하게 다른 의미로 지시하지 않는 이상, 복수의 표현을 포함할 수 있으며, 유사하게 복수로 표현되어 있다고 하더라도 단수의 의미를 포함할 수 있음을 알아야 한다.

[0027] 본 명세서의 전체에 걸쳐서 어떤 구성 요소가 다른 구성 요소를 "포함"한다고 기재하는 경우에는, 특별히 반대되는 의미의 기재가 없는 한 임의의 다른 구성 요소를 제외하는 것이 아니라 임의의 다른 구성 요소를 더 포함할 수도 있다는 것을 의미할 수 있다.

[0028] 더 나아가서, 어떤 구성 요소가 다른 구성 요소의 "내부에 존재하거나, 연결되어 설치된다"고 기재한 경우에는, 이 구성 요소가 다른 구성 요소와 직접적으로 연결되어 있거나 접촉하여 설치되어 있을 수 있고, 일정한 거리를 두고 이격되어 설치되어 있을 수도 있으며, 일정한 거리를 두고 이격되어 설치되어 있는 경우에 대해서는 해당 구성 요소를 다른 구성 요소에 고정 내지 연결시키기 위한 제 3의 구성 요소 또는 수단이 존재할 수 있으며, 이 제 3의 구성 요소 또는 수단에 대한 설명은 생략될 수도 있음을 알아야 한다.

[0029] 반면에, 어떤 구성 요소가 다른 구성 요소에 "직접 연결"되어 있다거나, 또는 "직접 접속"되어 있다고 기재되는 경우에는, 제 3의 구성 요소 또는 수단이 존재하지 않는 것으로 이해하여야 한다.

[0030] 마찬가지로, 각 구성 요소 간의 관계를 설명하는 다른 표현들, 즉 " ~ 사이에"와 "바로 ~ 사이에", 또는 " ~ 에 이웃하는"과 " ~ 에 직접 이웃하는" 등도 마찬가지의 취지를 가지고 있는 것으로 해석되어야 한다.

[0031] 또한, 본 명세서에 있어서 "일면", "타면", "일측", "타측", "제 1", "제 2" 등의 용어는, 사용된다면, 하나의 구성 요소에 대해서 이 하나의 구성 요소가 다른 구성 요소로부터 명확하게 구별될 수 있도록 하기 위해서 사용되며, 이와 같은 용어에 의해서 해당 구성 요소의 의미가 제한적으로 사용되는 것은 아님을 알아야 한다.

[0032] 또한, 본 명세서에서 "상", "하", "좌", "우" 등의 위치와 관련된 용어는, 사용된다면, 해당 구성 요소에 대해서 해당 도면에서의 상대적인 위치를 나타내고 있는 것으로 이해하여야 하며, 이들의 위치에 대해서 절대적인 위치를 특정하지 않는 이상은, 이들 위치 관련 용어가 절대적인 위치를 언급하고 있는 것으로 이해하여서는 아니된다.

[0033] 더욱이, 본 발명의 명세서에서는, "…부", "…기", "모듈", "장치" 등의 용어는, 사용된다면, 하나 이상의 기능이나 동작을 처리할 수 있는 단위를 의미하며, 이는 하드웨어 또는 소프트웨어, 또는 하드웨어와 소프트웨어의 결합으로 구현될 수 있음을 알아야 한다.

[0034] 또한, 본 명세서에서는 각 도면의 각 구성 요소에 대해서 그 도면 부호를 명기함에 있어서, 동일한 구성 요소에 대해서는 이 구성 요소가 비록 다른 도면에 표시되더라도 동일한 도면 부호를 가지고 있도록, 즉 명세서 전체에 걸쳐 동일한 참조 부호는 동일한 구성 요소를 지시하고 있다.

[0035] 본 명세서에 첨부된 도면에서 본 발명을 구성하는 각 구성 요소의 크기, 위치, 결합 관계 등은 본 발명의 사상을 충분히 명확하게 전달할 수 있도록 하기 위해서 또는 설명의 편의를 위해서 일부 과장 또는 축소되거나 생략되어 기술되어 있을 수 있고, 따라서 그 비례나 축척은 엄밀하지 않을 수 있다.

[0036] 또한, 이하에서, 본 발명을 설명함에 있어서, 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 구성, 예를 들어, 종래 기술을 포함하는 공지 기술에 대한 상세한 설명은 생략될 수도 있다.

[0037] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미가 있다.

[0039] 인공 지능(artificial intelligence, AI)은 인간의 지능으로 할 수 있는 사고, 학습, 자기계발 등을 컴퓨터가

할 수 있도록 하는 방법을 연구하는 컴퓨터 공학 및 정보기술의 한 분야로, 컴퓨터가 인간의 지능적인 행동을 모방할 수 있도록 하는 것을 의미한다.

[0040] 또한, 인공지능은 그 자체로 존재하는 것이 아니라, 컴퓨터 과학의 다른 분야와 직간접으로 많은 관련을 맺고 있다. 특히 현대에는 정보기술의 여러 분야에서 인공지능적 요소를 도입하여, 그 분야의 문제 풀이에 활용하려는 시도가 매우 활발하게 이루어지고 있다.

[0041] 머신 러닝(machine learning)은 인공지능의 한 분야로, 컴퓨터에 명시적인 프로그램 없이 배울 수 있는 능력을 부여하는 연구 분야이다.

[0042] 구체적으로 머신 러닝은, 경험적 데이터를 기반으로 학습을 하고 예측을 수행하고 스스로의 성능을 향상시키는 시스템과 이를 위한 알고리즘을 연구하고 구축하는 기술이라 할 수 있다. 머신 러닝의 알고리즘들은 엄격하게 정해진 정적인 프로그램 명령들을 수행하는 것이라기보다, 입력 데이터를 기반으로 예측이나 결정을 이끌어내기 위해 특정한 모델을 구축하는 방식을 취한다.

[0043] 인공신경망은 생물학적 뉴런의 동작원리와 뉴런간의 연결 관계를 모델링한 것으로 노드(node) 또는 처리 요소(processing element)라고 하는 다수의 뉴런들이 레이어(layer) 구조의 형태로 연결된 정보처리 시스템이다.

[0044] 인공 신경망은 기계 학습에서 사용되는 모델로써, 기계학습과 인지과학에서 생물학의 신경망(동물의 중추신경계 중 특히 뇌)에서 영감을 얻은 통계학적 학습 알고리즘이다.

[0045] 구체적으로 인공신경망은 시냅스(synapse)의 결합으로 네트워크를 형성한 인공 뉴런(노드)이 학습을 통해 시냅스의 결합 세기를 변화시켜, 문제 해결 능력을 가지는 모델 전반을 의미할 수 있다.

[0046] 용어 인공신경망은 용어 뉴럴 네트워크(Neural Network)와 혼용되어 사용될 수 있다.

[0047] 인공신경망은 복수의 레이어(layer)를 포함할 수 있고, 레이어들 각각은 복수의 뉴런(neuron)을 포함할 수 있다. 또한 인공신경망은 뉴런과 뉴런을 연결하는 시냅스를 포함할 수 있다.

[0048] 인공 신경망은 일반적으로 다음의 세가지 인자, 즉 (1) 다른 레이어의 뉴런들 사이의 연결 패턴 (2) 연결의 가중치를 갱신하는 학습 과정 (3) 이전 레이어로부터 수신되는 입력에 대한 가중 합으로부터 출력값을 생성하는 활성화 함수에 의해 정의될 수 있다.

[0049] 인공 신경망은, DNN(Deep Neural Network), RNN(Recurrent Neural Network), BRDNN(Bidirectional Recurrent Deep Neural Network), MLP(Multilayer Perceptron), CNN(Convolutional Neural Network)와 같은 방식의 네트워크 모델들을 포함할 수 있으나, 이에 한정되지 않는다.

[0050] 본 명세서에서 용어 '레이어'는 용어 '계층'과 혼용되어 사용될 수 있다.

[0051] 인공신경망은 계층 수에 따라 단층 신경망(Single-Layer Neural Networks)과 다층 신경망(Multi-Layer Neural Networks)으로 구분된다.

[0052] 일반적인 단층 신경망은, 입력층과 출력층으로 구성된다.

[0053] 또한 일반적인 다층 신경망은 입력층(Input Layer)과 하나 이상의 은닉층(Hidden Layer), 출력층(Output Layer)으로 구성된다.

[0054] 입력층은 외부의 자료들을 받아들이는 층으로서, 입력층의 뉴런 수는 입력되는 변수의 수와 동일하며, 은닉층은 입력층과 출력층 사이에 위치하며 입력층으로부터 신호를 받아 특성을 추출하여 출력층으로 전달한다. 출력층은 은닉층으로부터 신호를 받고, 수신한 신호에 기반한 출력 값을 출력한다. 뉴런간의 입력신호는 각각의 연결강도(가중치)와 곱해진 후 합산되며 이 합이 뉴런의 임계치보다 크면 뉴런이 활성화되어 활성화 함수를 통하여 획득한 출력값을 출력한다.

[0055] 한편 입력층과 출력 층 사이에 복수의 은닉층을 포함하는 심층 신경망은, 기계 학습 기술의 한 종류인 딥 러닝을 구현하는 대표적인 인공 신경망일 수 있다.

[0056] 한편 용어 '딥 러닝'은 용어 '심층 학습'과 혼용되어 사용될 수 있다.

[0057] 인공 신경망은 훈련 데이터(training data)를 이용하여 학습(training)될 수 있다. 여기서 학습이란, 입력 데이터를 분류(classification)하거나 회귀분석(regression)하거나 군집화(clustering)하는 등의 목적을 달성하기 위하여, 학습 데이터를 이용하여 인공 신경망의 파라미터(parameter)를 결정하는 과정을 의미할 수 있다. 인공

신경망의 파라미터의 대표적인 예시로써, 시냅스에 부여되는 가중치(weight)나 뉴런에 적용되는 편향(bias)을 들 수 있다.

[0058] 훈련 데이터에 의하여 학습된 인공 신경망은, 입력 데이터를 입력 데이터가 가지는 패턴에 따라 분류하거나 군집화 할 수 있다.

[0059] 한편 훈련 데이터를 이용하여 학습된 인공 신경망을, 본 명세서에서는 학습 모델(a trained model)이라 명칭 할 수 있다.

[0060] 다음은 인공 신경망의 학습 방식에 대하여 설명한다.

[0061] 인공 신경망의 학습 방식은 크게, 지도 학습, 비 지도 학습, 준 지도 학습(Semi-Supervised Learning), 강화 학습(Reinforcement Learning)으로 분류될 수 있다.

[0062] 지도 학습은 훈련 데이터로부터 하나의 함수를 유추해내기 위한 기계 학습의 한 방법이다.

[0063] 그리고 이렇게 유추되는 함수 중, 연속 적인 값을 출력하는 것을 회귀분석(Regression)이라 하고, 입력 벡터의 클래스(class)를 예측하여 출력하는 것을 분류(Classification)라고 할 수 있다.

[0064] 지도 학습에서는, 훈련 데이터에 대한 레이블(label)이 주어진 상태에서 인공 신경망을 학습시킨다.

[0065] 여기서 레이블이란, 훈련 데이터가 인공 신경망에 입력되는 경우 인공 신경망이 추론해 내야 하는 정답(또는 결과 값)을 의미할 수 있다.

[0066] 본 명세서에서는 훈련 데이터가 입력되는 경우 인공 신경망이 추론해 내야 하는 정답(또는 결과값)을 레이블 또는 레이블링 데이터(labeling data)이라 명칭 한다.

[0067] 또한 본 명세서에서는, 인공 신경망의 학습을 위하여 훈련 데이터에 레이블을 설정하는 것을, 훈련 데이터에 레이블링 데이터를 레이블링(labeling) 한다고 명칭한다.

[0068] 이 경우 훈련 데이터와 훈련 데이터에 대응하는 레이블은 하나의 트레이닝 셋(training set)을 구성하고, 인공 신경망에는 트레이닝 셋의 형태로 입력될 수 있다.

[0069] 한편 훈련 데이터는 복수의 특징(feature)을 나타내고, 훈련 데이터에 레이블이 레이블링 된다는 것은 훈련 데이터가 나타내는 특징에 레이블이 달린다는 것을 의미할 수 있다. 이 경우 훈련 데이터는 입력 객체의 특징을 벡터 형태로 나타낼 수 있다.

[0070] 인공 신경망은 훈련 데이터와 레이블링 데이터를 이용하여, 훈련 데이터와 레이블링 데이터의 연관 관계에 대한 함수를 유추할 수 있다. 그리고, 인공 신경망에서 유추된 함수에 대한 평가를 통해 인공 신경망의 파라미터가 결정(최적화)될 수 있다.

[0071] 생성적 적대 신경망이란, 생성자(generator)와 판별자(discriminator), 두 개의 서로 다른 인공지능이 경쟁하며 성능을 개선하는 머신 러닝 방법이다.

[0072] 이 경우 생성자는 새로운 데이터를 창조하는 모형으로, 원본 데이터를 기반으로 새로운 데이터를 생성할 수 있다.

[0073] 또한 판별자는 데이터의 패턴을 인식하는 모형으로, 입력된 데이터가 원본 데이터인지 또는 생성자에서 생성한 새로운 데이터인지 여부를 감별하는 역할을 수행할 수 있다.

[0074] 그리고 생성자는 판별자를 속이지 못한 데이터를 입력 받아 학습하며, 판별자는 생성자로부터 속은 데이터를 입력 받아 학습할 수 있다. 이에 따라 생성자는 판별자를 최대한 잘 속이도록 진화할 수 있고, 판별자는 원본 데이터와 생성자에 의해 생성된 데이터를 잘 구분하도록 진화할 수 있다.

[0076] 인공 신경망은 모델의 구성, 활성 함수(Activation Function), 손실 함수(Loss Function) 또는 비용 함수(Cost Function), 학습 알고리즘, 최적화 알고리즘 등에 의해 그 구조가 특정되며, 학습 전에 하이퍼파라미터(Hyperparameter)가 미리 설정되고, 이후에 학습을 통해 모델 파라미터(Model Parameter)가 설정되어 내용이 특정될 수 있다.

[0077] 예컨대, 인공 신경망의 구조를 결정하는 요소에는 은닉층의 개수, 각 은닉층에 포함된 은닉 노드의 개수, 입력

특징 벡터(Input Feature Vector), 대상 특징 벡터(Target Feature Vector) 등이 포함될 수 있다.

[0078] 하이퍼파라미터는 모델 파라미터의 초기값 등과 같이 학습을 위하여 초기에 설정하여야 하는 여러 파라미터들을 포함한다. 그리고, 모델 파라미터는 학습을 통하여 결정하고자 하는 여러 파라미터들을 포함한다.

[0079] 예컨대, 하이퍼파라미터에는 노드 간 가중치 초기값, 노드 간 편향 초기값, 미니 배치(Mini-batch) 크기, 학습 반복 횟수, 학습률(Learning Rate) 등이 포함될 수 있다. 그리고, 모델 파라미터에는 노드 간 가중치, 노드 간 편향 등이 포함될 수 있다.

[0080] 손실 함수는 인공 신경망의 학습 과정에서 최적의 모델 파라미터를 결정하기 위한 지표(기준)로 이용될 수 있다. 인공 신경망에서 학습은 손실 함수를 줄이기 위하여 모델 파라미터들을 조작하는 과정을 의미하며, 학습의 목적은 손실 함수를 최소화하는 모델 파라미터를 결정하는 것으로 볼 수 있다.

[0081] 손실 함수는 주로 평균 제곱 오차(MSE: Mean Squared Error) 또는 교차 엔트로피 오차(CEE, Cross Entropy Error)를 사용할 수 있으며, 본 발명이 이에 한정되지는 않는다.

[0082] 교차 엔트로피 오차는 정답 레이블이 원 핫 인코딩(one-hot encoding)된 경우에 사용될 수 있다. 원 핫 인코딩은 정답에 해당하는 뉴런에 대하여만 정답 레이블 값을 1로, 정답이 아닌 뉴런은 정답 레이블 값이 0으로 설정하는 인코딩 방법이다.

[0083] 머신 러닝 또는 딥 러닝에서는 손실 함수를 최소화하기 위하여 학습 최적화 알고리즘을 이용할 수 있으며, 학습 최적화 알고리즘에는 경사 하강법(GD: Gradient Descent), 확률적 경사 하강법(SGD: Stochastic Gradient Descent), 모멘텀(Momentum), NAG(Nesterov Accelerate Gradient), Adagrad, AdaDelta, RMSProp, Adam, Nadam 등이 있다.

[0084] 경사 하강법은 현재 상태에서 손실 함수의 기울기를 고려하여 손실 함수값을 줄이는 방향으로 모델 파라미터를 조정하는 기법이다.

[0085] 모델 파라미터를 조정하는 방향은 스텝(step) 방향, 조정하는 크기는 스텝 사이즈(size)라고 칭한다.

[0086] 이때, 스텝 사이즈는 학습률을 의미할 수 있다.

[0087] 경사 하강법은 손실 함수를 각 모델 파라미터들로 편미분하여 기울기를 획득하고, 모델 파라미터들을 획득한 기울기 방향으로 학습률만큼 변경하여 갱신할 수 있다.

[0088] 확률적 경사 하강법은 학습 데이터를 미니 배치로 나누고, 각 미니 배치마다 경사 하강법을 수행하여 경사 하강의 빈도를 높인 기법이다.

[0089] Adagrad, AdaDelta 및 RMSProp는 SGD에서 스텝 사이즈를 조절하여 최적화 정확도를 높이는 기법이다. SGD에서 모멘텀 및 NAG는 스텝 방향을 조절하여 최적화 정확도를 높이는 기법이다. Adam은 모멘텀과 RMSProp를 조합하여 스텝 사이즈와 스텝 방향을 조절하여 최적화 정확도를 높이는 기법이다. Nadam은 NAG와 RMSProp를 조합하여 스텝 사이즈와 스텝 방향을 조절하여 최적화 정확도를 높이는 기법이다.

[0090] 인공 신경망의 학습 속도와 정확도는 인공 신경망의 구조와 학습 최적화 알고리즘의 종류뿐만 아니라, 하이퍼파라미터에 크게 좌우되는 특징이 있다. 따라서, 좋은 학습 모델을 획득하기 위하여는 적당한 인공 신경망의 구조와 학습 알고리즘을 결정하는 것뿐만 아니라, 적당한 하이퍼파라미터를 설정하는 것이 중요하다.

[0091] 통상적으로 하이퍼파라미터는 실험적으로 다양한 값으로 설정해가며 인공 신경망을 학습시켜보고, 학습 결과 안정적인 학습 속도와 정확도를 제공하는 최적의 값으로 설정한다.

[0093] 이하 첨부된 도면들을 참조하여 본 발명의 일 실시 예에 따른 비지도 학습 기반의 깊이 예측 방법 및 시스템에 대해 설명하기로 한다.

[0094] 도 1은 본 발명의 일 실시 예에 따른 비지도 학습 기반의 깊이 측정 시스템의 네트워크 예시도이다.

[0095] 도 1을 참조하면, 본 발명의 일 실시 예에 따른 비지도 학습 기반의 깊이 예측 시스템(100)은, 네트워크(400)를 통해 서버(200) 및 사용자 단말(300)과 통신 연결될 수 있다.

[0096] 하드웨어에 해당하는 서버(200)는 각종 소프트웨어 서버, 예를 들어 클라우드 서버, 웹서버, 데이터베이스 서버, 인공지능 API 서버 및 데이터 마이닝 서버를 포함하도록 구성될 수 있다. 그리고 비지도 학습 기반의 깊

이 예측 시스템(100)은 서버(200)와의 통신을 통해 서버가 제공하는 각종 서비스를 제공받을 수 있다.

[0097] 사용자 단말(300)은, 서버(200) 및 비지도 학습 기반의 깊이 예측 시스템(100)에 통신 연결할 수 있는 클라이언트에 해당한다. 사용자는 다양한 모바일 기기, 예를 들어 스마트 폰, 태블릿 컴퓨터를 통해 비지도 학습 기반의 깊이 예측 시스템(100)에 접속하여, 본 발명의 일 실시 예에 따른 비지도 학습 기반의 깊이 예측 방법을 실행할 수 있다.

[0098] 네트워크(400)는 유선 및 무선 네트워크, 예를 들어 LAN(local area network), WAN(wide area network), 인터넷(internet), 인트라넷(intranet) 및 엑스트라넷(extranet), 그리고 모바일 네트워크, 예를 들어 셀룰러, 3G, LTE, WiFi 네트워크, 애드혹 네트워크 및 이들의 조합을 비롯한 임의의 적절한 통신 네트워크 일 수 있다.

[0099] 네트워크(400)는 허브, 브리지, 라우터, 스위치 및 게이트웨이와 같은 네트워크 요소들의 연결을 포함할 수 있다. 네트워크(400)는 인터넷과 같은 공용 네트워크 및 안전한 기업 사설 네트워크와 같은 사설 네트워크를 비롯한 하나 이상의 연결된 네트워크들, 예컨대 다중 네트워크 환경을 포함할 수 있다. 네트워크(400)에의 액세스는 하나 이상의 유선 또는 무선 액세스 네트워크들을 통해 제공될 수 있다.

[0100] 도 2는 본 발명의 일 실시 예에 따른 비지도 학습 기반의 깊이 측정 시스템의 블록도이다.

[0101] 도 2를 참조하면, 본 발명의 일 실시 예에 따른 비지도 학습 기반의 깊이 예측 시스템(100)은, 제어부(110), 입력 장치(120), 출력 장치(130), 저장 장치(140), 통신 장치(150), 및 메모리(160)를 포함하도록 구성될 수 있다. 깊이 측정 시스템의 핵심이 되는, 인공 신경망, 신경망(170) 또는 네트워크는, 학습부(161), 손실 함수 설정부(162), 파라미터 설정부(163) 및 평가부(164)는 하드웨어로 구현될 수도 있으나, 도 2와 같이 소프트웨어 형태로 구현되어서, 메모리(160)에 저장되어 있으면서 제어부(110)의 호출에 따라 구동될 수 있다.

[0102] 제어부(110)는, 프로세서(processor)를 포함하도록 구성될 수 있다. 제어부(110)는 입력 장치(120), 출력 장치(130), 저장 장치(140), 통신 장치(150) 및 메모리(160)의 동작을 제어하는 기능을 수행한다. 특히 제어부(110)는, 본 발명과 관련하여 신경망(170)의 학습, 테스트, 평가를 수행하고, 평가를 마친 신경망을 이용하여 최종적으로 2D의 싱글 이미지를 3D 이미지로 변환하는 기능을 수행할 수 있다.

[0103] 입력 장치(120), 출력 장치(130), 저장 장치(140), 통신 장치(150)는 컴퓨팅 장치, 즉 컴퓨터의 구성에서 키보드, 마우스, 터치 패드, 디스플레이, SSD, HDD, 이더넷 카드, 및 무선랜 모듈 등을 포함하여, 미래 선보일 다양한 형태의 장치로 구현될 수 있다.

[0104] 학습부(161)는 본 발명의 일 실시 예에 따른 신경망(170), 즉 GAN 신경망 기반의 깊이 예측 네트워크를 학습시키는 기능을 수행한다.

[0105] 손실 함수 설정부(162)는 학습에 사용될 훈련 손실 함수를 정의하고, 그 범위를 결정하는 기능을 수행한다.

[0106] 파라미터 설정부(163)는 학습 과정에서 학습에 사용되는 각종 데이터셋 관련 파라미터를 설정하는 기능을 수행한다.

[0107] 평가부(164)는, 학습을 마친 신경망을 대상으로 깊이 예측의 테스트 및 평가 과정을 수행할 수 있다.

[0108] 신경망(170)은, 본 발명의 일 실시 예에서 인공 신경망 또는 네트워크로 지칭될 수 있으며, 다양한 종류의 신경망 중에서 생성적 적대적 네트워크(Generative Adversarial Network, GAN)를 주로 포함하도록 구성될 수 있다.

[0110] 도 3은 본 발명의 일 실시 예에 따른 비지도 학습 기반의 깊이 측정 방법의 흐름도이다.

[0111] 도 3을 참조하면, 본 발명의 일 실시 예에 따른 비지도 학습 기반의 깊이 측정 방법(S100)은, 학습 데이터셋을 이용하는 신경망 학습(S110), 학습된 신경망의 테스트(S120), 신경망 테스트 결과의 평가(S130) 및 신경망을 이용하여 이미지의 깊이 예측(S140)을 포함하도록 구성될 수 있다.

[0113] 도 4는 본 발명의 일 실시 예에 따른 비지도 학습 기반의 깊이 측정 방법의 흐름도이다.

[0114] 도 4를 참조하면, 비지도 학습 기반의 깊이 측정 방법(S100) 중에서, 학습 데이터셋을 이용하는 신경망 학습(S100)은, S210 내지 S240 단계를 포함하도록 구성될 수 있다.

[0115] S210 단계에서 비지도 학습 기반의 깊이 측정 시스템(100)은, 좌측 이미지를 입력 받거나, 또는 수신할 수

있다.

[0116] S220 단계에서 비지도 학습 기반의 깊이 측정 시스템(100)은, 입력 받은 좌측 이미지 기반으로 2개의 시차맵을 생성할 수 있다.

[0117] S230 단계에서 비지도 학습 기반의 깊이 측정 시스템(100)은, 좌측 이미지와 시차맵을 이용하여 제1 쌍의 페이크 이미지를 생성할 수 있다.

[0118] S240 단계에서 비지도 학습 기반의 깊이 측정 시스템(100)은, 대칭 손실을 포함하는 훈련 손실 함수를 이용하여 판별자에 의한 실제 및 페이크 이미지를 판별할 수 있다.

[0119] 또한, 비지도 학습 기반의 깊이 측정 시스템(100)은, 훈련 손실 함수에 포함된 대칭 손실에 관한 학습을 위해 다음의 관계를 위의 과정과는 따로 또는 병행할 수 있다.

[0120] S215 단계에서 비지도 학습 기반의 깊이 측정 시스템(100)은, 우측 이미지를 입력 받거나, 또는 수신할 수 있다.

[0121] S225 단계에서 비지도 학습 기반의 깊이 측정 시스템(100)은, 입력 받은 우측 이미지 기반으로 2개의 시차맵을 생성할 수 있다.

[0122] S235 단계에서 비지도 학습 기반의 깊이 측정 시스템(100)은, 우측 이미지와 시차맵을 이용하여 제2 쌍의 페이크 이미지를 생성할 수 있다. 제2 쌍의 페이크 이미지는 제1 쌍의 페이크 이미지와 함께, 깊이 예측 신경망의 대칭 손실의 학습에 이용될 수 있다.

[0123] 이하 도 5 내지 도 7을 참조하여, 비지도 학습 기반의 깊이 측정 시스템(100)에 포함된 깊이 측정 신경망(170)의 동작에 대해 자세히 설명하기로 한다.

[0124] 본 발명의 일 실시 예에 따른 깊이 예측 네트워크 및 훈련 손실에 대해 설명한다. 본 발명의 해결 과제는 단일 컬러 이미지 I가 입력으로 주어졌을 때 픽셀 단위 시차맵(disparity map) 또는 역 깊이 map을 예측하는 생성자(d=G(I))를 훈련시키는 것이다. 깊이 p는 예측된 disparity, d와 p=(b · f)/d의 관계를 갖는다. 여기서 b는 두 카메라 사이의 기준 거리, f는 카메라 초점 거리를 의미한다. 본 발명의 일 실시 예에서 제안하는 비지도 또는 자가 감독(self-supervised) 방식의 적대적 깊이 예측 네트워크는 이미지 재구성 손실을 점진적으로 최소화하는 방향으로 훈련된다. 깊이 정답 데이터의 지도 없이 오직 스테레오 이미지 쌍만 사용하여 네트워크가 훈련된다. 대칭적 형태의 생성자 네트워크, 상대론적 판별자 네트워크, 그리고 본 발명의 일 실시 예에 따른 손실함수에 대해서 설명하기로 한다.

[0126] 도 5는 본 발명의 일 실시 예에 따른 단안 영상의 깊이 예측을 위해 제안된 비지도 학습 기반의 GAN 신경망의 예시도이다.

[0127] 도 5를 참조하면, 본 발명의 일 실시 예에 따른 GAN 신경망(170)은 메인 네트워크(175) 및 보조 네트워크(176)를 포함하도록 구성될 수 있다. 메인 네트워크는 이미지의 생성 및 판별과 연관된다. 보조 네트워크(176)은 이미지의 생성과 연관된다.

[0128] 도 5에서 좌측 이미지와 우측 이미지는 상호 호환적이다. 상호 호환적 2가지 실시 예 중에서 메인 네트워크(175)에 좌측 이미지가 입력되고, 보조 네트워크(176)에 우측 이미지가 입력되는 실시 예를 대표로 하여 설명한다.

[0129] 메인 네트워크(175)는, 제1 생성자(171), 제1 판별자(172) 및 제2 판별자(173)를 포함하도록 구성될 수 있다.

[0130] 제1 생성자(171)는, 좌측 이미지($I_l$)를 입력 받아, 좌측 이미지 기반의 시차맵 2개($d_r^{(1)}$, $d_l^{(1)}$)를 생성한다. 메인 네트워크(175)는 시차맵과 좌측 이미지 및 우측 이미지를 이용하여 제1 쌍의 페이크 이미지, 즉 좌측 이미지 기반의 우측 페이크 이미지($\hat{I}_r^{(1)}$)와 좌측 페이크 이미지($\hat{I}_l^{(1)}$)를 생성한다. 좌측 페이크 이미지는 좌측 이미지, 즉 좌측 실제 이미지와 비교되어 제2 판별자(173)에 의해 판별된다. 그리고, 우측 페이크 이미지는 우측 이미지, 즉 우측 실제 이미지와 비교되어 제1 판별자(172)에 의해 판별된다.

[0131] 보조 네트워크(176)는 제2 생성자(174)를 포함하도록 구성될 수 있다.

[0132] 보조 네트워크(176) 및 제2 생성자(174)는, 메인 네트워크(175)가 좌측 이미지에 대해 수행한 과정과 동일한 과정을 우측 이미지($I_r$)에 대해 수행할 수 있다. 그 결과 보조 네트워크(176)은 제2 쌍의 페이크 이미지, 즉 우측 이미지 기반의 우측 페이크 이미지($\hat{I}_r^{(r)}$)와 좌측 페이크 이미지($\hat{I}_l^{(r)}$)를 생성할 수 있다. 제2 쌍의 페이크 이미지는 제1 쌍의 페이크 이미지와 함께 대칭적 손실의 학습에 이용될 수 있다.

[0133] 도 5는 본 발명에서 제안하는 자가 감독 방식의 적대적 깊이 예측 네트워크 프레임워크를 나타낸다. 네트워크는 크게 메인 네트워크와 서브 네트워크로 구성된다. 본 발명의 훈련 대상인 메인 네트워크는 하나의 생성자($G_l$) 와 두 개의 판별자($D_r$, $D_l$)로 이루어져 있다. 메인 네트워크의 학습 과정은 다음과 같다.

[0135] 도 6은 본 발명의 일 실시 예에 따른 예측된 시차맵을 이용한 이미지 재구성의 예시도이다.

[0136] 도 6과 같이 Gl은 왼쪽 이미지($I_l$)를 입력받아 두 개의 disparity map, $d_r^{(l)}$, $d_l^{(l)}$을 생성한다. 그 다음 Il을 $d_r^{(l)}$에 따라 재구성하여 페이크 우측 이미지($\hat{I}_r^{(l)}$)를 만든다. 같은 방식으로 실제 오른쪽 이미지($I_r$)를 $d_l^{(l)}$에 따라 재구성하여 가짜 왼쪽 이미지 $\hat{I}_l^{(l)}$를 생성한다. 이후 판별자 네트워크가 생성된 가짜 이미지와 실제 이미지를 입력받아 가짜가 실제보다 얼마나 덜 현실적인지, 진짜가 가짜보다 얼마나 더 현실적인지에 대한 확률을 리턴한다.

[0137] 서브 네트워크는 하나의 생성자 네트워크($G_r$)만으로 이루어져 있다. Gr은 Ir을 입력 받아 두 개의 disparity map $d_r^{(r)}$, $d_l^{(r)}$을 생성한다. 마찬가지로 두 개의 가짜 이미지가 생성된다. 본 발명은 이들과 메인 네트워크의 가짜 이미지의 L1 손실을 계산하여 대칭적 손실(symmetric loss)로 정의한다. $G_r$은 $G_l$의 현재상태의 복사본으로써 의미적으로만 구분되며, 대칭적 손실 이외에 다른 손실에는 전혀 관여하지 않는다. 본 발명은 이를 통해 오른쪽 이미지에 대한 간접적인 학습이 가능하다고 보았다. 결과적으로 왼쪽-오른쪽 일관성 (left-right consistency)을 더욱 뚜렷하게 함으로써 occlusion 및 dis-occlusion 문제를 개선할 수 있다고 간주한다.

[0138] 도 5는 또한 훈련을 위한 4개의 주요 손실함수를 보여준다. 첫 번째로 판별자와의 경쟁을 통한 성능 개선을 위한 적대적 손실(adversarial loss)이 있다. 두 번째로 이미지 재구성 손실(image reconstruction loss)은 실제 이미지와 재구성된 이미지의 차이를 나타낸다. 본 발명은 이를 위해 L1 손실과 GMSD IQA 알고리즘을 결합하였다. 세 번째로 사전 훈련된 네트워크 모델의 관점에서 두 이미지가 얼마나 다른지 평가하여 훈련에 반영하기 위한 인지적 손실(perceptual loss)이 있다. 마지막으로 위에서 언급한 대칭적 손실 (symmetric loss)을 통해 합리적인 추가 훈련 오버헤드로 좌-우 일관성을 향상시키고 disparity map의 정확성을 더욱 견고하게 한다.

[0139] 생성적 적대적 신경망(Generative Adversarial, Network, GAN)은 이미지 합성 분야에서 뛰어난 성능을 보이고 있지만 일반적으로 다음과 같은 심각한 문제를 안고 있다. 첫째, 모델 매개 변수가 진동하고 불안정해지면서 수렴하지 않는다. 둘째, 생성자가 붕괴되어 제한된 종류의 샘플이 생성된다. 셋째, 판별자의 학습이 빨라 생성자의 기울기가 소실되면서 생성자의 학습이 멈춘다. 넷째, 생성자와 판별자의 불균형이 과적합(over-fitting)을 유발한다. 다섯째, 하이퍼 매개 변수 선택에 매우 민감하다. 이러한 문제들을 해결하기 위해 LSGAN, WGAN-GP와 같은 방법론이 제안되었다.

[0140] 종래 기술에 따르면 일반적인 GAN이 갖지 못한 주요 속성은 가짜 데이터가 실제일 확률이 높아질수록 실제 데이터가 실제일 확률이 낮아진다는 것이다. 본 발명은 이를 만족하기 위해 상대론적 평균 GAN의 개념을 차용한다. 일반적인 판별자 D는 입력 이미지 I가 실제 또는 가짜($D(I)=\sigma(I)$)일 가능성을 추정하는 반면 상대론적 판별자는 실제 이미지 $I_R$이 가짜 이미지 $I_F$ 보다 얼마나 더 현실적인지 그리고 $I_F$, $I_R$ 보다 얼마나 덜 현실적일 확률 ($D(I_R, I_F)= \sigma(C(I_R)-E_{IF}[C(I_F)])$)을 예측한다. 여기서 C(x)는 변형되지 않은 판별자의 출력, σ는 시그모이드(sigmoid) 함수, 그리고 $E_{IR/F}[\cdot]$는 미니 배치 내 모든 진짜 또는 가짜 이미지들의 평균을 취하는 것을 나타낸다. 판별자의 손실 $L_D$는 다음과 같이 정의된다.

*수학식 1*

$$L_{D_r} = -\mathbb{E}_{I_r}[\log(D_r(I_r, \hat{I}_r^{(l)}))] - \mathbb{E}_{\hat{I}_r^{(l)}}[1-\log(D_r(\hat{I}_r^{(l)}, I_r))]$$

$$L_{D_l} = -\mathbb{E}_{I_l}[\log(D_l(I_l, \hat{I}_l^{(l)}))] - \mathbb{E}_{\hat{I}_l^{(l)}}[1-\log(D_l(\hat{I}_l^{(l)}, I_l))]$$

$$L_D = L_{D_r} + L_{D_l}$$

[0141]

[0143] 생성자를 위한 적대적 손실 $L_{adv}$는 다음과 같이 정의된다.

*수학식 2*

$$L_{adv_r} = -\mathbb{E}_{I_r}[1-log(D_r(I_r, \hat{I}_r^{(l)})] - \mathbb{E}_{\hat{I}_r^{(l)}}[log(D_r(\hat{I}_r^{(l)}, I_r)]$$

[0144]

*수학식 3*

$$L_{adv_l} = -\mathbb{E}_{I_l}[1 - log(D_l(I_l, \hat{I}_l^{(l)})] - \mathbb{E}_{\hat{I}_l^{(l)}}[log(D_l(\hat{I}_l^{(l)}, I_l)]$$

[0146]

*수학식 4*

$$L_{adv} = L_{adv_r} + L_{adv_l}$$

[0147]

[0149] 본 발명은 실제에 가까운 가짜 이미지로 재구성하기 위한 중간 단계로써 네트워크 모델을 훈련시킨다. 즉, 학습 네트워크는 더 나은 품질의 가짜 이미지를 생성하기 위한 disparity map을 예측하기 위한 방법을 훈련 받는다. 깊이 정답 데이터의 지도 없이 스테레오 이미지 쌍을 이용하므로 이미지 재구성 손실을 최소화해야 한다. 따라서 두 이미지 간의 품질 차이를 보다 정확하고 빠르게 가려내는 것이 중요하다. 종래에는 이러한 목적으로 대부분 L1 손실과 SSIM IQA 알고리즘을 혼합하는 방식을 주로 사용했다. 본 발명은 이에 영감을 받아 SSIM을 더 나은 성능의 IQA 알고리즘으로 대체하여 결과 모델의 성능을 더욱 향상시킬 수 있다고 간주하고, 이미지 재구성 손실 함수의 일부로 GMSD(gradient magnitude similarity deviation)을 사용한다. 먼저 실제 이미지와 재구성된 이미지 사이의 픽셀 거리 절대값을 최소화하기 위한 L1 손실 (L1이 L2 손실보다 뛰어난 성능을 보임)을 다음과 같이 정의한다.

*수학식 5*

$$L_{1_r} = \| I_r - f_{w+}(d_r^{(l)}, I_l) \|_1 = \| I_r - \hat{I}_r^{(l)} \|_1$$

$$L_{1_l} = \| I_l - f_{w+}(d_l^{(l)}, I_r) \|_1 = \| I_l - \hat{I}_l^{(l)} \|_1$$

[0150]

[0152] 일반적으로 SSIM, MS-SSIM, CW-SSIM 및 FSIM과 같은 IQA 알고리즘은 지역 품질 맵(regional quality map)의 평

균 또는 가중 평균으로부터 최종 품질 점수는 얻는다. 그러나 종래 기술은 지역 품질 맵의 표준 편차에서 더욱 정확한 인지적 이미지 품질을 얻기 위해 표준 편차 풀링 전략을 제안했다. 그들은 이미지 그라디언트가 왜곡에 민감하고 왜곡된 이미지 내 다양한 지역적 구조들이 서로 다른 정도(degree)로 영향을 받는다는 사실에 주목하였다. 그들은 원본 이미지의 그라디언트 크기와 비왜곡된 이미지를 지역적으로 비교하여 GMS (gradient magnitude similarity) 맵을 생성한다. 품질이 좋지 않은 이미지는 일반적으로 로컬 품질 변화가 크기 때문에 GMS 맵의 표준 편차 값이 최종 품질 점수로 결정된다. 그들은 Prewitt 연산을 사용하여 그라디언트를 계산하였다. 그들이 제시한 GMS map 계산식은 다음과 같다.

*수학식 6*

[0153]
$$GMS(i)=\frac{2m_r(i)\cdot m_d(i)+c}{m_r^2(i)+m_d^2(i)+c}$$

[0155] 여기서 c는 수치적 안정성을 위한 양의 상수이며, $m_r(i)$와 $m_d(i)$는 각각 원본 이미지와 왜곡된 가짜 이미지의 그라디언트 크기를 나타낸다. 그들은 또한 GMS의 표준 편차를 사용하여 최종 품질 점수를 얻기 위해 다음 식을 정의하였다.

*수학식 7*

[0156]
$$GMSM=\frac{1}{N}\sum_{i=1}^{N}GMS(i)$$

$$GMSD=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(GMS(i)-GMSM)^2}$$

[0158] 여기서 N은 이미지의 총 픽셀 개수이다. 제시된 기법에 따라 본 발명의 이미지 재구성 손실 함수는 다음과 같이 정의된다(실험에서 $\alpha=0.85$)

*수학식 8*

[0159]
$$L_{rec_r}=\alpha\cdot L_{L1_r}+(1-\alpha)\cdot GMSD(I_r,\hat{I}_r^{(l)})$$

*수학식 9*

[0161]
$$L_{rec_l}=\alpha\cdot L_{L1_l}+(1-\alpha)\cdot GMSD(I_l,\hat{I}_l^{(l)})$$

*수학식 10*

[0163]
$$L_{rec}=L_{rec_r}+L_{rec_l}$$

[0165] GMSD를 손실 함수로 사용하는 것은 여전히 전통적인 영상 처리 기법에 의존한다고 볼 수 있다. 본 발명은 사전 훈련된 네트워크 모델의 feature map을 통해 이를 보완하고자 한다. 인지적 손실은 이미지 간 콘텐츠나 스타일과 같은 높은 수준의 차이를 비교하는데 쓰이는 손실 함수로써 주로 초고해상도 이미지 변환 등에 사용된다. 본 발명은 이것이 실제 이미지와 재구성된 이미지 사이에도 적용 가능하며 궁극적으로 재구성된 이미지의 품질을 향상시키는데 기여할 것으로 가정한다. ESRGAN에서 영감을 받아 사전 학습되고 미세 조정된(fine-tuned) VGG-19 모델을 인지적 손실 계산에 사용한다.v(x)를 이미지 x를 입력받는 사전 훈련된 VGG-19 네트워크 모델이라고 할 때 인지적 손실은 다음과 같이 정의 된다.

*수학식 11*

[0166] $$L_{percept_l} = ||v(I_r) - v(\hat{I}_r^{(l)})||_1$$

*수학식 12*

[0168] $$L_{percept_r} = ||v(I_l) - v(\hat{I}_l^{(l)})||_1$$

*수학식 13*

[0169] $$L_{percept} = L_{percept_l} + L_{percept_r}$$

[0171] 본 발명은 단안 영상의 깊이 예측을 위해 두 개의 생성자를 사용한다. 메인 네트워크의 생성자 $G_l$은 왼쪽 이미지를 입력받아 두 개의 disparity map을 생성한다. 반대로 서브 네트워크의 생성자 $G_r$은 오른쪽 이미지를 입력 받는다. 본 발명의 목표는 $G_l$을 훈련시키는 것이며, $G_r$은 $G_l$을 훈련시키기 위한 손실 함수처럼 동작한다. 이렇게 $G_r$을 사용하는 이유는 다음과 같다. 먼저 훈련을 위해 두 쌍의 이미지가 주어지며 올바른 뷰 이미지에 대해 배우지 않는 것은 낭비라고 할 수 있다. 오른쪽 이미지를 학습하면 학습 데이터가 더 풍부해져 모델의 성능을 향상시킬 수 있다. 그러나 $G_l$과 동일하게 $G_r$을 훈련하면 오른쪽 이미지에 대한 모든 손실을 계산해야 한다. 결과적으로 훈련 시간이 두 배가 된다. 따라서 우리는 현재 모델에 오른쪽 이미지를 입력하여 얻은 결과를 훈련을 위한 손실 데이터로 사용한다. 서브 네트워크에서 생성된 가짜 이미지들은 대칭적 손실이라고 하는 메인 네트워크의 가짜 이미지들과 각각 비교된다. 본 발명은 대칭적 손실을 통해 이미지의 좌/우 경계 영역과 occlusion 및 dis-occlusion 문제를 개선할 수 있다고 가정했다. 또한 네트워크 모델이 좌/우 일관성을 유지하는 데 유용하다. 대칭적 손실 은 두 이미지의 L1 손실로 정의된다.

*수학식 14*

$$L_{sym_r} = \| \hat{I}_r^l - \hat{I}_r^r \|_1$$

$$L_{sym_l} = \| \hat{I}_l^l - \hat{I}_l^r \|_1$$

$$L_{sym} = L_{sym_r} + L_{sym_l}$$

[0172]

[0173] 종래 기술대로 좌-우 일관성 손실을 다음과 같이 정의한다. 대칭적 손실 은 서로 다른 생성자 네트워크로부터 파생된 가짜 이미지를 비교하지만, 좌-우 일관성 손실은 메인 네트워크의 disparity map으로 제한된다.

*수학식 15*

$$d_{left-right} = f_{w_+}(d_l^{(l)}, d_r^{(l)})$$

$$d_{right-left} = f_{w_-}(d_r^{(l)}, d_l^{(l)})$$

$$L_{con_r} = \| d_{left-right} - d_r^{(l)} \|_1$$

$$L_{con_l} = \| d_{right-left} - d_l^{(l)} \|_1$$

$$L_{con} = L_{con_r} + L_{con_l}$$

[0174]

[0175] 또한 본 발명은 edge aware smoothness 손실 $L_{smooth}$을 사용한다.

[0176] 지금까지 제안하는 적대적 깊이 예측 모델에 대한 훈련 손실 함수를 정의했다. 각 손실 함수는 최적화를 위한 가중치를 가지며, 전체 손실은 다음과 같이 정의된다.

*수학식 16*

[0177] $$L_{total} = \lambda_1 L_{adv} + \lambda_2 L_{rec} + \lambda_3 L_{percept} + \lambda_4 L_{sym} + \lambda_5 L_{con} + \lambda_6 L_{smooth}$$

[0179] 본 발명의 생성자 네트워크는 종래 기술로부터 영감을 받아 ResNet50 백본 네트워크를 인코더 파트로 사용한다. 디코더 파트는 ELU 연산 계층을 포함하는 5개의 디콘볼루션으로 구성된다. 또한 이중 선형 샘플링(bilinear sampling)을 통해 2배 크기로 업샘플링(up-sampling)을 수행한다. 본 발명은 아티팩트 생성을 줄이기 위해, 손실을 계산할 때 다운샘플링(down-sampling)된 다중 스케일 이미지들과 생성된 disparity map을 원래 해상도로 복원한다. 경계 아티팩트를 줄이기 위해 제로 패딩(zero padding) 대신 반사 패딩(reflection padding)을 사용한다.

[0181] 도 7은 본 발명의 일 실시 예에 따른 판별자 네트워크의 블록도이다.

[0182] 여기서, k, n 및 s는 각각 커널 크기, 특징 map 개수 및 스트라이드를 나타낸다. 판별자는 중간 단계에서 컨볼

루션 레이어 이후 배치 정규화(batch normalization), 드롭 아웃(drop out), 리키 렐루(leaky ReLU)를 수행하고, 이는 샘플링 데이터의 무작위성을 높이고 안정적인 적대적 훈련을 가능하게 한다.

[0184] 한편 상술한 본 발명의 설명에서는 구체적인 실시 예에 관해 설명하였으나, 여러 가지 변형이 본 발명의 범위에서 벗어나지 않고 실시될 수 있다.

[0185] 전술한 본 발명은, 프로그램이 기록된 매체에 컴퓨터가 읽을 수 있는 코드로서 구현하는 것이 가능하다. 컴퓨터가 읽을 수 있는 매체는, 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다. 컴퓨터가 읽을 수 있는 매체의 예로는, HDD(Hard Disk Drive), SSD(Solid State Disk), SDD(Silicon Disk Drive), ROM, RAM, CD-ROM, 자기 테이프, 플로피 디스크, 광 데이터 저장 장치 등이 있으며, 또한 캐리어 웨이브(예를 들어, 인터넷을 통한 전송)의 형태로 구현되는 것도 포함한다. 또한, 컴퓨터는 제어부(110)를 포함할 수도 있다.

[0186] 한편 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 다양한 수정 및 변형이 가능할 것이다. 따라서 상기의 상세한 설명은 모든 면에서 제한적으로 해석되어서는 아니 되고 예시적인 것으로 고려되어야 한다. 본 발명의 범위는 첨부된 청구항의 합리적 해석에 의해 결정되어야 하고, 본 발명의 등가적 범위 내에서의 모든 변경은 본 발명의 범위에 포함된다.

[0188] 본 발명의 일 실시 예에 따르면, 향상된 생성적 적대적 네트워크를 이용하여 단안 영상으로부터 고품질의 양안의 스테레오 이미지를 획득할 수 있다.

[0189] 또한, 각종 특징적인 생성적 적대적 네트워크의 손실 함수를 이용하여 실제 영상에 근접한 품질의 가짜 이미지가 생성될 수 있다.

[0190] 또한, 합리적인 훈련 오버헤드와 우측 이미지에 대한 간접적인 학습을 위한 대칭적 손실을 통해 예측되는 시차맵의 정확성이 향상될 수 있다.

[0192] 이상, 일부 예를 들어서 본 발명의 바람직한 여러 가지 실시 예에 대해서 설명하였지만, 본 "발명을 실시하기 위한 구체적인 내용" 항목에 기재된 여러 가지 다양한 실시 예에 관한 설명은 예시적인 것에 불과한 것이며, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자라면 이상의 설명으로부터 본 발명을 다양하게 변형하여 실시하거나 본 발명과 균등한 실시를 행할 수 있다는 점을 잘 이해하고 있을 것이다.

[0193] 또한, 본 발명은 다른 다양한 형태로 구현될 수 있기 때문에 본 발명은 상술한 설명에 의해서 한정되는 것이 아니며, 이상의 설명은 본 발명의 개시 내용이 완전해지도록 하기 위한 것으로 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에게 본 발명의 범주를 완전하게 알려주기 위해 제공되는 것일 뿐이며, 본 발명은 청구범위의 각 청구항에 의해서 정의될 뿐임을 알아야 한다.

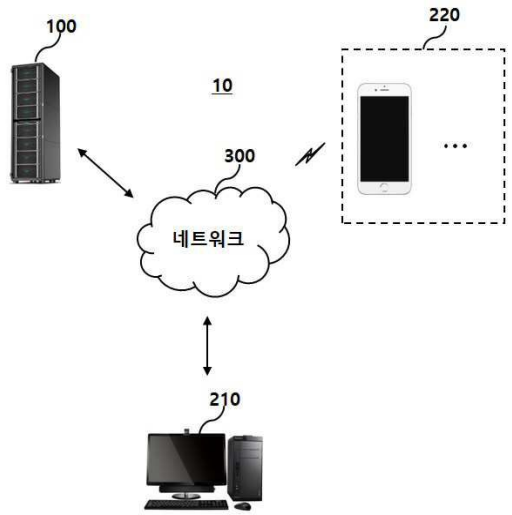
## *부호의 설명*

[0194] 100: 깊이 예측 시스템

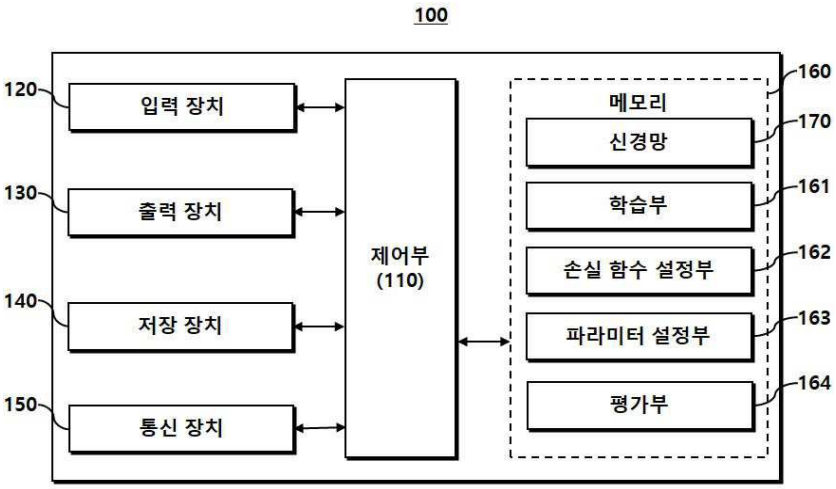
170: 깊이 예측 신경망

도면

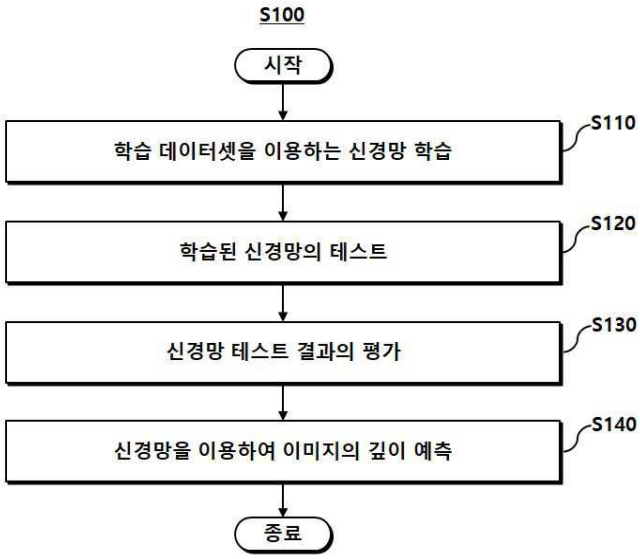
도면1


100
10
220
300
네트워크
210

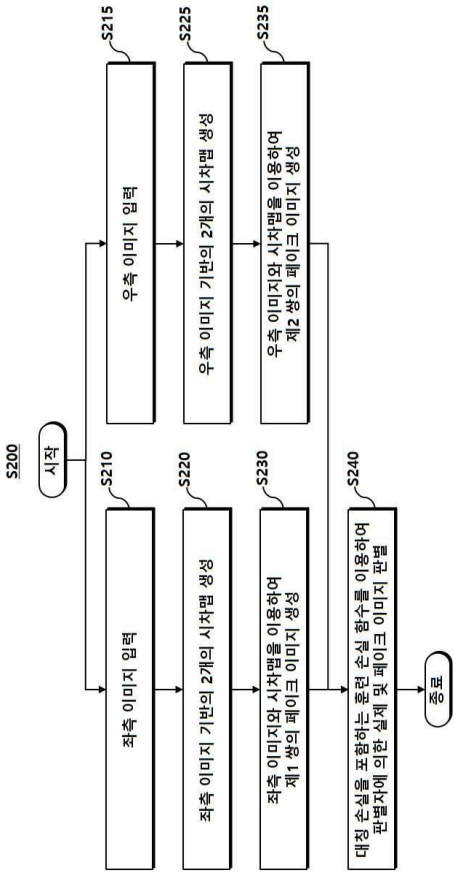
도면2


100
120
입력 장치
130
출력 장치
140
저장 장치
150
통신 장치
제어부
(110)
메모리
160
신경망
170
학습부
161
손실 함수 설정부
162
파라미터 설정부
163
평가부
164

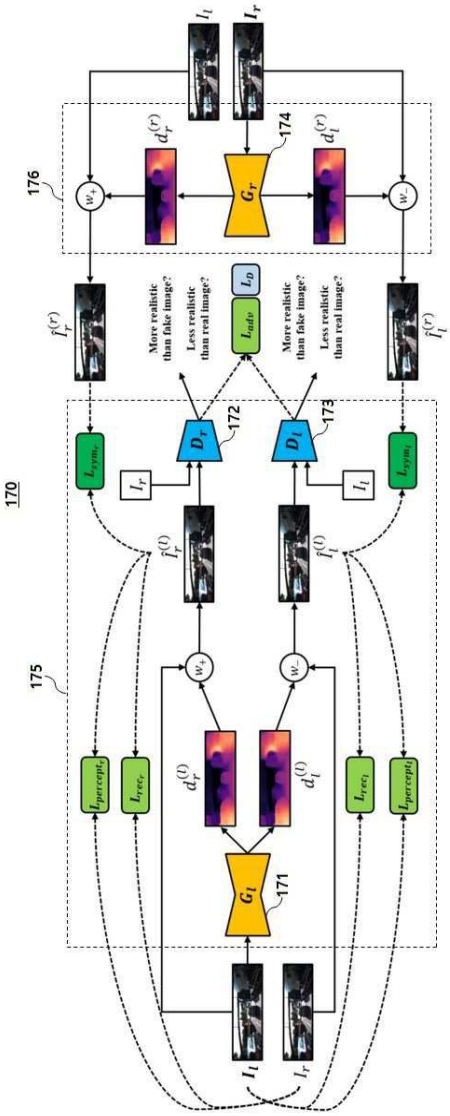
도면3


S100
시작
학습 데이터셋을 이용하는 신경망 학습
S110
학습된 신경망의 테스트
S120
신경망 테스트 결과의 평가
S130
신경망을 이용하여 이미지의 깊이 예측
S140
종료

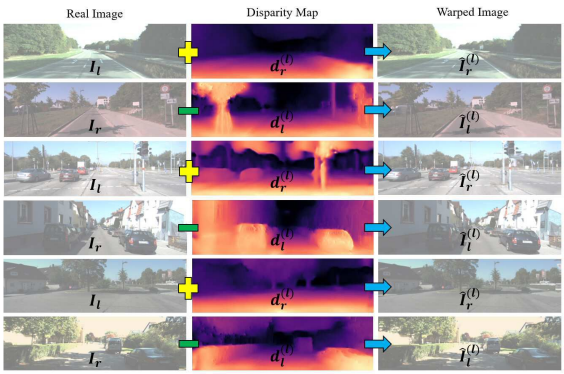
도면4


S200
시작
S210
좌측 이미지 입력
S215
우측 이미지 입력
S220
좌측 이미지 기반의 2개의 시차맵 생성
S225
우측 이미지 기반의 2개의 시차맵 생성
S230
좌측 이미지와 시차맵을 이용하여 제1 쌍의 페이크 이미지 생성
S235
우측 이미지와 시차맵을 이용하여 제2 쌍의 페이크 이미지 생성
S240
대칭 손실을 포함하는 훈련 손실 함수를 이용하여 판별자에 의한 실제 및 페이크 이미지 판별
종료

도면5


176
174
170
172
173
175
171
More realistic than fake image?
Less realistic than real image?
More realistic than fake image?
Less realistic than real image?

도면6


Real Image
Disparity Map
Warped Image

도면7

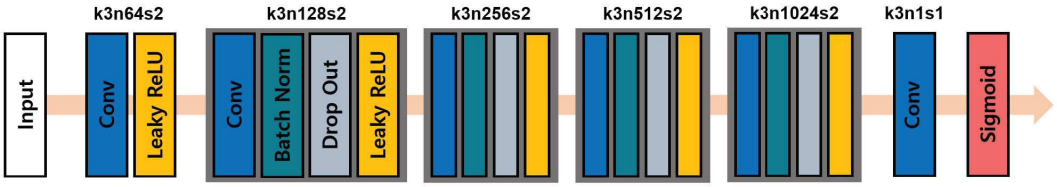



FIGURE 3. Illustration of the proposed discriminator network