



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2023년01월11일  
(11) 등록번호 10-2487270  
(24) 등록일자 2023년01월06일

(51) 국제특허분류(Int. Cl.)  
G06T 7/593 (2017.01) G06N 3/04 (2023.01)  
G06N 3/08 (2023.01)  
(52) CPC특허분류  
G06T 7/593 (2017.01)  
G06N 3/045 (2023.01)  
(21) 출원번호 10-2019-7010331  
(22) 출원일자(국제) 2017년09월12일  
심사청구일자 2020년08월25일  
(85) 번역문제출일자 2019년04월10일  
(65) 공개번호 10-2019-0065287  
(43) 공개일자 2019년06월11일  
(86) 국제출원번호 PCT/GB2017/052671  
(87) 국제공개번호 WO 2018/046964  
국제공개일자 2018년03월15일  
(30) 우선권주장  
1615470.0 2016년09월12일 영국(GB)  
(56) 선행기술조사문헌  
Ravi Garg 등, Unsupervised CNN for Single  
View Depth Estimation: Geometry to the  
Rescue, arXiv:1603.04992v1.(2016.05.16.)\*  
(뒷면에 계속)

(73) 특허권자  
나이앤틱, 인크.  
미국, 캘리포니아주 94111, 샌프란시스코, 스위트  
200, 원 페리 빌딩  
(72) 발명자  
고다드, 클레멘트  
미국, 캘리포니아주 94111, 샌프란시스코, 스위트  
200, 원 페리 빌딩, 나이앤틱, 인크. 사내  
맥 아오다, 오이신  
미국, 캘리포니아주 94111, 샌프란시스코, 스위트  
200, 원 페리 빌딩, 나이앤틱, 인크. 사내  
브로스토프, 가브리엘  
미국, 캘리포니아주 94111, 샌프란시스코, 스위트  
200, 원 페리 빌딩, 나이앤틱, 인크. 사내  
(74) 대리인  
김영철, 김 순 영

전체 청구항 수 : 총 20 항

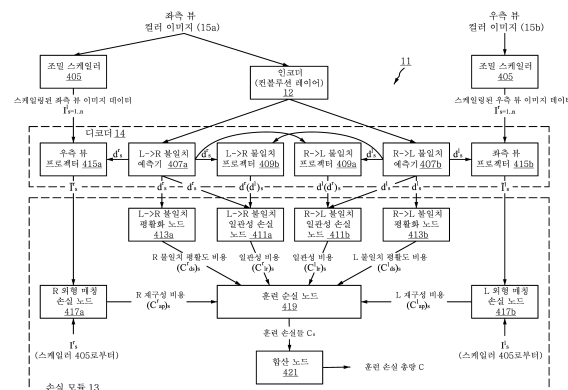
심사관 : 이재원

(54) 발명의 명칭 통계적 모델을 이용한 이미지 데이터로부터의 깊이 예측

(57) 요약

시스템 및 방법은 컨볼루션 뉴럴 네트워크(CNN)와 같은 통계 모델을 사용하여 컬러 이미지 데이터로부터 깊이를 예측하기 위해 기술된다. 모델은 양안 스테레오 이미지 쌍에 대해 학습되고, 단일 소스 컬러 이미지로부터 깊이 데이터를 예측할 수 있다. 입력 양안 스테레오 쌍의 각 이미지에 대해, 모델은 이미지에 적용될 때 다른 이미지의 재구성을 가능하게 하는 대응하는 불일치 값을 예측하도록 훈련된다. 모델은 스테레오 쌍의 각 이미지에 대한 예측된 불일치 값들 사이의 일관성을 강화하는 비용 함수에 기초하여 업데이트된다.

대표도



(52) CPC특허분류

**G06N 3/084** (2023.01)

G06T 2207/10024 (2013.01)

G06T 2207/20076 (2013.01)

G06T 2207/20081 (2013.01)

G06T 2207/20084 (2013.01)

(56) 선행기술조사문헌

Wenqiao Zhu 등, Variational Stereo Matching with Left Right Consistency Constraint, 2011 SoCPaR.(2011.10.14.)\*

JP2011525657 A

KR1020150043546 A

KR1020160056132 A

\*는 심사관에 의하여 인용된 문헌

---

## 명세서

### 청구범위

#### 청구항 1

단일 입력 컬러 이미지로부터 불일치 값을 예측하기 위한 모델을 제공하는 단계;

상기 모델을 이용하여 단일 입력 컬러 이미지로부터 불일치 값을 생성하는 단계; 및

상기 단일 입력 컬러 이미지로부터 생성된 불일치 값을 이용하여 상기 단일 입력 컬러 이미지에 대응하는 추정 값이 데이터를 산출하는 단계;를 포함하고,

상기 모델은 좌측 이미지 및 우측 이미지를 포함하는 적어도 한쌍의 입력 양안 스테레오 이미지 쌍에 대해

상기 좌측 이미지 또는 상기 우측 이미지 중 하나만 사용하여, 좌측-우측 불일치 값 및 우측-좌측 불일치 값을 예측하는 단계 - 상기 좌측-우측 불일치 값은 상기 좌측 이미지에 적용될 때 예측된 우측 이미지의 재구성을 가능하게 하고, 상기 우측-좌측 불일치 값은 상기 우측 이미지에 적용될 때 예측된 좌측 이미지의 재구성을 가능하게 함 -; 및

상기 예측된 좌측-우측 및 우측-좌측 불일치 값들 사이의 일관성을 강화하는 비용 함수에 기초하여 상기 모델을 업데이트하는 단계에 의하여, 훈련되는 컴퓨터로-구현되는 방법.

#### 청구항 2

제1 항에 있어서,

상기 모델은

상기 예측된 좌측-우측 불일치 값을 이용하여 투영된 우측-좌측 불일치 값을 계산하는 단계; 및

상기 예측된 우측-좌측 불일치 값을 이용하여 투영된 좌측-우측 불일치 값을 계산하는 단계;를 더 포함하여 훈련되는 컴퓨터로-구현되는 방법.

#### 청구항 3

제2 항에 있어서,

상기 투영된 우측-좌측 불일치 값은 예측된 좌측-우측 불일치 값을 샘플링하고, 상기 샘플링된 데이터에 다른 이미지의 예측된 우측-좌측 불일치 값을 적용함으로써 계산되고, 그리고

상기 투영된 좌측-우측 불일치 값은 예측된 우측-좌측 불일치 값을 샘플링하고, 상기 샘플링된 데이터에 다른 이미지의 예측된 좌측-우측 불일치 값을 적용함으로써 계산되는 컴퓨터로-구현되는 방법.

#### 청구항 4

제2 항에 있어서,

상기 비용 함수는 예측된 좌측-우측 및 우측-좌측 불일치 값과 투영된 좌측-우측 및 우측-좌측 불일치 값 사이의 일관성을 강화하는 불일치 일관성 컴포넌트를 포함하는 컴퓨터로-구현되는 방법.

#### 청구항 5

제1 항에 있어서,

상기 모델은

예측된 좌측-우측 불일치 값을 적용하여 상기 좌측 이미지의 샘플링된 픽셀을 쉬프트함으로써 상기 우측 이미지를 재구성하는 단계; 및

예측된 우측-좌측 불일치 값을 적용하여 상기 우측 이미지의 샘플링된 픽셀을 쉬프트함으로써 상기 좌측 이미지를 재구성하는 단계;를 더 포함하여 훈련되는 컴퓨터로-구현되는 방법.

## 청구항 6

제3 항에 있어서,

상기 샘플링은 이중선형 보간법(bilinear interpolation)을 포함하는 컴퓨터로-구현되는 방법.

## 청구항 7

제5 항에 있어서,

상기 비용 함수는 예측된 좌측 및 우측 이미지의 재구성과 스테레오 쌍의 각각의 좌측 및 우측 이미지 사이의 이미지 재구성 오차를 최소화하는 재구성된 외형 매칭 컴포넌트를 더 포함하는 컴퓨터로-구현되는 방법.

## 청구항 8

제7 항에 있어서,

상기 비용 함수는 상기 예측된 좌측-우측 및 우측-좌측 불일치 값에서 국부적인 평활화를 장려하는 평활도 컴포넌트를 더 포함하는 컴퓨터로-구현되는 방법.

## 청구항 9

제8 항에 있어서,

상기 비용 함수는 불일치 일관성 컴포넌트, 상기 평활도 컴포넌트, 및 상기 재구성된 외형 매칭 컴포넌트의 가중 합을 구현하는 컴퓨터로-구현되는 방법.

## 청구항 10

제1 항에 있어서,

상기 모델은 컨볼루션 뉴럴 네트워크를 포함하고,

상기 컨볼루션 뉴럴 네트워크는 처리 노드들의 구조화된 배열을 포함하며, 각각의 처리 노드는 적어도 하나의 가중치를 가지는 컴퓨터로-구현되는 방법.

## 청구항 11

제10 항에 있어서,

상기 컨볼루션 뉴럴 네트워크는 상기 비용 함수의 역-전파 컴포넌트에 의해 훈련되는 컴퓨터로-구현되는 방법.

## 청구항 12

제1 항에 있어서,

상기 모델은

복수의 공간 해상도에서 상기 입력 양안 스테레오 이미지 쌍의 상기 좌측 및 우측 이미지를 업-샘플링하고 업-컨볼루션하는 단계; 및

각 공간 해상도에서 각각의 좌측-우측 및 우측-좌측 불일치 값을 예측하는 단계를 더 포함하여 훈련되고,

상기 모델은 상기 스테레오 이미지 쌍의 각각의 이미지에 대한 각 공간 해상도에서 상기 예측된 좌측-우측 및 우측-좌측 불일치 값들 사이의 일관성을 강화하는 비용 함수에 기초하여 업데이트되었던 것인 컴퓨터로-구현되는 방법.

## 청구항 13

제12 항에 있어서,

상기 비용 함수는 상기 공간 해상도에 의존하는 상기 예측된 좌측-우측 및 우측-좌측 불일치 값들 사이의 일관성의 가중된 강화를 포함하는 컴퓨터로-구현되는 방법.

#### 청구항 14

제1 항에 있어서,

상기 양안 스테레오 이미지 쌍은 공지된 카메라 초점 거리 및 공지된 베이스라인 거리를 두고 각각의 카메라에 의해 동시에 캡처되고, 상기 양안 스테레오 이미지 쌍에 따라 대응하는 깊이 데이터가 예측된 좌측-우측 및 우측-좌측 불일치 값으로부터 계산되는 컴퓨터로-구현되는 방법.

#### 청구항 15

제14 항에 있어서,

상기 양안 스테레오 이미지 쌍은 정렬되고 시간적으로 정렬된 스테레오 쌍인 컴퓨터로-구현되는 방법.

#### 청구항 16

제15 항에 있어서,

디지털 이미지는 상기 이미지를 캡처한 각각의 카메라의 속성을 정의하는 메타 데이터로 해석되는 컴퓨터로-구현되는 방법.

#### 청구항 17

단일 입력 컬러 이미지로부터 불일치 값을 예측하기 위한 모델을 정의하는 데이터를 저장하는 단계; 및

좌측 이미지 및 우측 이미지를 포함하는 적어도 한쌍의 입력 양안 스테레오 이미지 쌍에 대해 모델을 훈련하는 단계로서, 상기 좌측 이미지 또는 상기 우측 이미지 중 하나만 사용하여, 좌측-우측 불일치 값 및 우측-좌측 불일치 값을 예측하고 - 상기 좌측-우측 불일치 값은 상기 좌측 이미지에 적용될 때 예측된 우측 이미지의 재구성을 가능하게 하고, 상기 우측-좌측 불일치 값은 상기 우측 이미지에 적용될 때 예측된 좌측 이미지의 재구성을 가능하게 함 -; 및 상기 예측된 좌측-우측 및 우측-좌측 불일치 값들 사이의 일관성을 강화하는 비용 함수에 기초하여 상기 모델을 업데이트함으로써, 상기 모델을 훈련하는 단계;를 포함하고,

훈련 모델은 단일 입력 컬러 이미지로부터 깊이 이미지를 생성하는데 사용될 수 있고, 상기 훈련 모델을 사용하여 상기 입력 컬러 이미지로부터 생성된 불일치 맵을 생성하고; 상기 단일 입력 컬러 이미지로부터 생성된 불일치 값을 이용하여 상기 단일 입력 컬러 이미지와 대응하는 추정된 깊이 데이터를 계산함으로써 상기 깊이 이미지를 생성하는 컴퓨터로-구현되는 방법.

#### 청구항 18

제17 항에 있어서,

상기 단일 입력 컬러 이미지는 카메라에 의해 캡처되는 컴퓨터로-구현되는 방법.

#### 청구항 19

명령을 저장하는 비 일시적 컴퓨터 판독 가능 매체로서, 상기 명령은 컴퓨팅 장치에 의해 실행될 때, 상기 컴퓨팅 장치가 단계들을 수행하도록 하며, 상기 단계들은,

단일 입력 컬러 이미지로부터 불일치 값을 예측하기 위한 모델을 제공하는 단계;

상기 모델을 이용하여 단일 입력 컬러 이미지로부터 불일치 값을 생성하는 단계; 및

상기 단일 입력 컬러 이미지로부터 생성된 불일치 값을 이용하여 상기 단일 입력 컬러 이미지에 대응하는 추정된 깊이 데이터를 산출하는 단계;를 포함하고,

상기 모델은 좌측 이미지 및 우측 이미지를 포함하는 적어도 한쌍의 입력 양안 스테레오 이미지 쌍에 대해

상기 좌측 이미지 또는 상기 우측 이미지 중 하나만 사용하여, 좌측-우측 불일치 값 및 우측-좌측 불일치 값을 예측하는 단계 - 상기 좌측-우측 불일치 값은 상기 좌측 이미지에 적용될 때 예측된 우측 이미지의 재구성을 가능하게 하고, 상기 우측-좌측 불일치 값은 상기 우측 이미지에 적용될 때 예측된 좌측 이미지의 재구성을 가능하게 함 -; 및

상기 예측된 좌측-우측 및 우측-좌측 불일치 값들 사이의 일관성을 강화하는 비용 함수에 기초하여 상기 모델을 업데이트하는 단계에 의하여, 훈련되는 비 일시적 컴퓨터 판독 가능 매체.

## 청구항 20

제19항에 있어서,

상기 단계는:

복수의 공간 해상도에서 상기 입력 양안 스테레오 이미지 쌍의 상기 좌측 및 우측 이미지를 업-샘플링하고 업-컨볼루션하는 단계; 및

각 공간 해상도에서 각각의 좌측-우측 및 우측-좌측 불일치 값을 예측하는 단계를 더 포함하고,

상기 모델은 상기 스테레오 이미지 쌍의 각각의 이미지에 대한 각 공간 해상도에서 상기 예측된 좌측-우측 및 우측-좌측 불일치 값들 사이의 일관성을 강화하는 비용 함수에 기초하여 업데이트되었던 것인 비 일시적 컴퓨터 판독 가능 매체.

## 발명의 설명

### 기술 분야

[0001] 본 발명은 이미지 데이터 처리 시스템에 관한 것으로, 보다 구체적으로 훈련된 통계적 모델을 이용하여 이미지 데이터로부터 깊이 데이터를 예측하는 이미지 데이터 처리 시스템에 관한 것이다.

### 배경 기술

[0002] 이미지로부터 깊이를 추정하는 것은 컴퓨터 비전(computer vision)에서 오랜 역사를 가진다. 유익한 접근 방식은 모션으로부터의 구조, X로부터의 형상, 양안 및 다안 스테레오에 의존해 왔다. 하지만, 이러한 기술의 대부분은 관심 있는 장면을 여러 번 관측할 수 있다는 과정에 의존하는 것이다. 이러한 관측은 다양한 조명 조건에서 장면을 관측하거나 여러 시점의 형태로 나타날 수 있다. 이러한 한계를 극복하기 위해, 최근 지도 학습 문제로서, 단 하나의 입력 이미지만 있는 단안 깊이 추정의 작업을 제기하는 연구, 예를 들어, 엘. 레디키, 제이. 샤이, 및 엠. 폴레프스, "원근법에서 어긋나는 물체 폴링", CVPR 2014. (L. Ladicky, J. Shi, and M. Pollefeys, "Pulling Things Out Of Perspective", CVPR 2014.), 디. 에이젠, 씨. 퍼스츠, 및 알. 퍼구스, "멀티-스케일 딥 네트워크를 사용한 단안 이미지로부터의 깊이 맵 추정", NIPS 2014. (D. Eigen, C. Puhrsch, and R. Fergus, "Depth Map Prediction From A Single Image Using A Multi-Scale Deep Network", NIPS 2014.) 및 에프. 리우, 씨. 쉐, 지. 린 및 아이. 레이드, "딥 컨볼루션 뉴럴 필드를 사용한 단일 모노컬러 이미지로부터 깊이 학습", PAMI 2015. (F. Liu, C. Shen, G. Lin, and I. Reid, "Learning Depth From Single Monocular Images Using Deep Convolutional Neural Fields", PAMI 2015.)에서 논의된 바와 같은 연구의 수가 증가하였다. 그러나, 이러한 연구에서 설명된 방법은 실제 측정 깊이 데이터(ground truth depth data)의 대규모 수집에 대해 오프라인 훈련 받은 모델을 사용하여 이미지의 각 픽셀의 깊이를 직접 예측하려고 한다. 따라서, 이러한 방법들은 대규모 이미지 수집과 해당 픽셀 깊이가 있는 장면으로 제한된다.

[0003] 훈련 동안 이미지 재구성 문제로서 자동 깊이 추정을 처리하는 대안적인 접근법이 개발되었다. 인간은 원근법, 익숙한 객체의 알려진 크기에 비례하는 스케일링, 조명 및 음영의 형태에 따른 외형, 어클루전(occlusion), 다른 단서 등과 같은 단서를 이용하여 단안의 깊이 추정을 잘 수행한다. 하향식 및 상향식 단서들의 조합은 전체 장면 이해와 깊이를 정확하게 추정하는 우리의 능력을 연결하는 것처럼 보인다. 최근, 공개된 몇몇의 논문들은 훈련 시간에 실제 측정 깊이(ground truth depth)를 필요로 하지 않는 새로운 뷰 합성과 깊이 추정을 위한 딥 네트워크 기반 방법을 제안한다.

[0004] 제이. 플린, 아이. 뉴랜더, 제이. 필빈, 및 엔. 스나벨리, "딥 스테레오: 세계의 이미지로부터 새로운 뷰를 예측하는 학습", CVPR 2016 (J. Flynn, I. Neulander, J. Philbin, and N. Snavely, "DeepStereo: Learning to Predict New Views from the World's Imagery", CVPR 2016.)는, 딥 스테레오(Deep stereo)라고 하는 새로운 이미지 합성 네트워크에 대해 논의한다. 딥 스테레오(Deep stereo)는 근처의 인접한 이미지로부터 픽셀을 선택하여 새로운 뷰를 생성한다. 훈련을 하는 동안, 그들은 이미지의 세트를 선택하고, 각각의 카메라의 자세 (모션

으로부터 움직임 측정(odometry) 및 표준 구조(standard structure)의 조합을 사용)를 계산한 다음, 주변 이미지의 외형을 예측하기 위해 CNN(convolutional neural network)을 훈련한다. 평면 스위프 볼륨(plane sweep volumes)을 기준으로 인접한 이미지의 색상을 샘플링 하는 데 가장 적합한 깊이가 선택된다. 테스트 시, 작은 오버랩 패치에 대해 이미지 합성이 수행된다. 그러나, 딥 스테레오(Deep stereo)는 테스트 시 주변 포즈 이미지가 여러 개 필요하기 때문에 단안 깊이 추정에 적합하지 않다.

[0005] 제이. 시에, 알. 거석, 및 에이. 파하디, "딥3d: 딥 컨볼루션 뉴럴 네트워크를 통한 완전 자동 2D에서 3D 비디오 변환", ECCV 2016. (J. Xie, R. Girshick, and A. Farhadi, "Deep3d: Fully Automatic 2D-To-3D Video Conversion With Deep Convolutional Neural Networks", ECCV 2016.) 또한 훈련 스테이지에서 새로운 뷰 합성의 문제를 다룬다. 이 스테이지의 목적은 양안 스테레오 이미지 쌍의 컨텍스트에서 입력 왼쪽 이미지(즉, 소스 이미지)로부터 대응하는 오른쪽 뷰를 생성하는 것이다. 컴퓨터 비전에서 잘 알려진 바와 같이, 양안 불일치(binocular disparity)는 카메라 사이의 수평 분리(시차)로 인해 발생하는 2 개의 스테레오 이미지들 내의 유사한 피쳐들의 좌표들, 즉 왼쪽 및 오른쪽 카메라들에 의해 보여진 물체의 이미지 위치의 차이에 해당한다. 딥 3D(Deep3D)는 양안 시차를 사용하여 입체 영상에서 2차원 이미지의 깊이 정보를 추출한다. 이미지 재구성 손실을 다시 사용하여, 그들의 방법은 입력 왼쪽 이미지의 각 픽셀에 대해 가능한 모든 불일치에 대한 분포를 생성한다. 최종 합성 된 오른쪽 이미지의 픽셀 값은 왼쪽 이미지와 동일한 스캔 라인상의 픽셀들의 조합이고, 각 불일치의 확률에 의해 가중된다. 그들의 이미지 형성 모델의 단점은 후보 불일치 값의 수가 증가하면 알고리즘의 메모리 소비가 크게 증가하므로, 큰 출력 해상도에 대한 접근 방식을 확장하는 것이 어렵다는 것이다.

[0006] 딥3D(Deep3D)와 유사하게, 알. 가그, 브이. 쿠마 비지, 및 아이. 레이드, "단일 뷰 깊이 예측을 위한 비지도 학습 CNN: 구조에 대한 지오메트리", ECCV 2016. (R. Garg, V. Kumar BG, and I. Reid, "Unsupervised CNN For Single View Depth Estimation: Geometry To The Rescue", ECCV 2016.)은 훈련 단계에서 양안 스테레오 이미지 쌍을 기반으로 한 이미지 재구성 손실을 사용하여 단안 깊이를 추정하는 CNN 훈련을 설명한다. 그러나, 가그 외.(Garg et al.) 에서 설명된 이미지 형성 모델은 완전히 미분 가능(fully differentiable)하지 않고, 부차적인(suboptimal) 훈련을 수행한다. 보상하기 위해, 가그 외.(Garg et al.)는 Taylor 근사를 수행하여 손실을 선형화하며, 결과적으로 최적화하기가 더 어려운 목표를 산출하게 된다.

[0007] 깊이 추정을 위한 상술한 deep CNN 기반 시스템의 한계를 해결하고 최종 결과의 품질을 크게 향상시키는 향상된 네트워크 아키텍처가 요구되고 있다.

## 발명의 내용

### 해결하려는 과제

[0008] 본 발명은 깊이 추정을 위한 상술한 deep CNN 기반 시스템의 한계를 해결하고 최종 결과의 품질을 크게 향상시키는 향상된 네트워크 아키텍처를 제공한다.

### 과제의 해결 수단

[0009] 본 발명의 양태는 첨부된 청구 범위에 기재되어있다.

[0010] 본 발명의 일 실시예는 컬러 이미지 데이터로부터 깊이 데이터를 예측하기 위한 통계 모델을 정의하는 데이터를 저장하는 단계; 및 적어도 한쌍의 입력 양안 스테레오 이미지 쌍에 대해 모델을 훈련하는 단계로서, 상기 입력 양안 스테레오 쌍의 각 이미지에 대해, 상기 이미지에 적용될 때 다른 이미지의 재구성을 가능하게 하는 대응하는 불일치 값을 예측하고; 상기 스테레오 쌍의 각 이미지에 대한 상기 예측된 불일치 값들 사이의 일관성을 강화하는 비용 함수에 기초하여 상기 모델을 업데이트함으로써, 상기 모델을 훈련하는 단계를 포함하는 컴퓨터 구현 방법을 제공한다.

[0011] 상기 모델을 훈련하는 단계는, 상기 스테레오 쌍의 각 이미지에 대해 상기 대응하는 불일치 값에 기초하여 투영된 불일치 값을 계산하는 단계를 더 포함할 수 있다. 상기 투영된 불일치 값은 상기 스테레오 쌍의 하나의 이미지에 대해, 제1 이미지의 예측된 불일치 값을 샘플링하고, 상기 샘플링된 데이터에 다른 이미지의 예측된 불일치 값을 적용함으로써 계산될 수 있다. 상기 비용 함수는 예측된 불일치 값과 스테레오 쌍의 각 이미지에 대해 계산된 투영된 불일치 값 사이의 일관성을 강화하는 불일치 일관성 컴포넌트를 포함할 수 있다.

[0012] 상기 모델의 재구성 모듈은 상기 대응하는 예측된 불일치 값을 적용하여 상기 스테레오 쌍의 제1 이미지의 샘플링된 이미지 픽셀을 쉬프트함으로써 상기 스테레오 쌍의 제2 이미지를 재구성할 수 있다. 상기 비용 함수는 상



기 재구성된 이미지와 대응하는 입력 이미지 사이의 이미지 재구성 오차를 최소화하는 재구성된 외형 매칭 컴포넌트를 더 포함할 수 있다. 샘플링은 이중선형 보간법을 포함할 수 있다.

[0013] 상기 비용 함수는 상기 대응하는 예측된 불일치 값에서 국부적인 평활화를 장려하는 평활도 컴포넌트를 더 포함할 수 있다. 상기 비용 함수는 상기 불일치 일관성 컴포넌트, 상기 평활도 컴포넌트, 및 상기 재구성된 외형 매칭 컴포넌트의 가중 합을 구현할 수 있다.

[0014] 상기 통계 모델은 컨볼루션 뉴럴 네트워크(CNN)를 포함할 수 있고, 상기 컨볼루션 뉴럴 네트워크(CNN)은 처리 노드들의 구조화된 배열을 포함하며, 상기 각각의 처리 노드는 적어도 하나의 파라미터 값을 가진다. 컨볼루션 뉴럴 네트워크(CNN)은 상기 비용 함수의 역-전파 컴포넌트에 의해 훈련될 수 있다.

[0015] 모델 훈련은 복수의 공간 해상도에서 상기 입력 이미지 데이터를 업-샘플링하고 업-컨볼루션하는 단계; 및 각 공간 해상도에서 대응하는 불일치 값을 예측하는 단계를 더 포함할 수 있고, 상기 모델은 상기 스테레오 쌍의 각각의 이미지에 대한 각 공간 해상도에서 상기 예측된 불일치 값들 사이의 일관성을 강화하는 비용 함수에 기초하여 업데이트된다. 상기 비용 함수는 상기 공간 해상도에 의존하는 상기 예측된 불일치 값들 사이의 일관성의 가중된 강화를 포함할 수 있다.

[0016] 상기 양안 스테레오 훈련 이미지 쌍은 공지된 카메라 초점 거리 및 공지된 베이스라인 거리를 두고 각각의 카메라에 의해 동시에 캡처될 수 있다. 상기 양안 스테레오 훈련 이미지 쌍은 정렬되고 시간적으로 정렬된 스테레오 쌍일 수 있다. 상기 디지털 이미지는 상기 이미지를 캡처한 각각의 카메라의 속성을 정의하는 메타 데이터로 해석될 수 있다.

[0017] 다른 실시예에 따라, 상기 훈련 모델을 사용하여 상기 입력 컬러 이미지로부터 예측된 불일치 맵을 생성하고; 상기 예측된 불일치 맵으로부터 대응하는 추정된 깊이 데이터를 계산함으로써 상기 깊이가 이미지가 생성될 수 있다. 상기 컬러 이미지 데이터는 카메라에 의해 캡처될 수 있다. 상기 모델은 큰 해상도 이미지를 수신하도록 구성될 수 있다.

[0018] 바람직하게는, 본 발명은 임의의 깊이 데이터를 요구하지 않는 완전 컨볼루션(convolutional) 모델을 제공하고 대신에 중간 단계로서 깊이를 합성하도록 훈련된다. 본 발명은 알려진 카메라 기준선을 갖는 정렬된 스테레오 이미지의 쌍 사이의 픽셀 레벨 대응성을 예측하도록 학습한다.

[0019] 또한, 실시예들은, 네트워크 내에서의 좌-우 불일치 일관성 제약을 포함하는 새로운 트레이닝 손실로 종단 간 지도되지 않은 단안 깊이 추정을 수행하는 네트워크 아키텍처; 몇몇 다른 훈련 손실 및 설명된 접근법의 효과를 강조하는 이미지 형성 모델에 대한 평가; 및 다른 상이한 데이터 세트에 일반화하는 모델을 포함한다.

[0020] 또 다른 실시예에 따르면, 본 발명은 단 하나의 입력 이미지만 있고, 장면 기하 구조 또는 존재하는 대상의 유형에 대한 가정이 없는 단안 깊이를 추정을 위한 지도되지 않은 딥 뉴럴 네트워크를 제공한다. 특정 구현 환경에서 사용 가능하지 않거나 획득하는 비용이 많이 들 수 있는 정렬된 실제 측정 깊이 데이터(ground truth depth data)를 사용하는 대신, 본 발명은 양안 스테레오 데이터가 캡처될 수 있는 용이성을 이용한다. 또 다른 실시예에 따르면, 학습 모듈은 훈련 중 각 카메라 뷰의 예측된 깊이 맵들 사이의 일관성을 강화하는 손실 함수를 구현하여 향상된 예측을 유도한다. 결과 출력 깊이는 훈련 단계에서 실제 측정 깊이 정보의 누락에도 불구하고, 완전히 지도된 기준선보다 우수하다. 또한, 훈련된 모델은 훈련 중에 보이지 않는 데이터 세트에 일반화할 수 있고 시각적으로 그럴듯한 깊이 맵을 생성할 수 있다.

[0021] 다른 실시예에서, 전술한 바와 같은 방법들을 수행하도록 구성된 장치 및 시스템들이 제공된다. 또 다른 실시예에서, 프로그래머블 장치로 하여금 전술한 바와 같은 방법을 수행하게 하는 기계 판독 가능한 명령을 포함하는 컴퓨터 프로그램이 제공된다.

## 발명의 효과

[0022] 본 발명은 단 하나의 입력 이미지만 있고, 장면 기하 구조 또는 존재하는 대상의 유형에 대한 가정이 없는 단안 깊이를 추정을 위한 지도되지 않은 딥 뉴럴 네트워크를 제공한다. 특정 구현 환경에서 사용 가능하지 않거나 획득하는 비용이 많이 들 수 있는 정렬된 실제 측정 깊이 데이터(ground truth depth data)를 사용하는 대신, 본 발명은 양안 스테레오 데이터가 캡처될 수 있는 용이성을 이용한다.

## 도면의 간단한 설명



- [0023] 아래에 설명된 도면을 참조하여, 단지 예시로서, 본 발명의 실시예에 대한 상세한 설명이 이루어진다.
- 도 1은 본 발명의 일 실시예에 따른 이미지 처리 시스템의 주요 구성을 도시한 블록도이다.
- 도 2는 예시적인 CNN의 섹션의 개략도이다.
- 도 3은, 도 3a 및 도 3b를 포함하고, 일 실시예에 따라 단일 이미지 깊이 예측 CNN을 훈련하기 위해 훈련 모듈에 의해 수행되는 주요 처리 단계를 도시한 순서도이다.
- 도 4는 일 실시예에 따른 반복 훈련에서 예시적인 CNN의 처리 및 데이터 구성 요소를 개략적으로 도시한 순서도이다.
- 도 5는 일 실시예에 따라 훈련된 CNN을 사용하여 단일 소스 이미지로부터 깊이 데이터를 생성 및 처리하는 예시적인 과정에 대한 순서도이다.
- 도 6은 실시예들의 하나 이상의 기능들이 구현될 수 있는 컴퓨터 시스템의 예시적인 도면이다.

### 발명을 실시하기 위한 구체적인 내용

- [0024] 도 1은 컬러 이미지 데이터로부터 깊이 데이터를 예측하고 처리하기 위한 예시적인 시스템(1)의 블록도이다. 도시된 바와 같이, 시스템(1)은 캡처된 뷰 내의 오브젝트를 형성하는 픽셀에 대한 RGB 값을 서술하는 RGB 이미지와 같은, 카메라(7)로부터 캡처된 컬러 이미지 데이터를 수신할 수 있는 깊이 데이터 생성 모듈(5)을 포함하는 이미지 처리 시스템(3)을 포함한다. 디지털 이미지는 이미지를 캡처한 각각의 카메라의 속성을 정의하는 메타데이터로 해석될 수 있다. 깊이 데이터 생성 모듈(5)은 훈련된 컨볼루션 뉴럴 네트워크(CNN) 모듈(11)의 불일치 예측기(9)를 사용하여 단일 소스 이미지의 수신된 컬러 이미지 데이터로부터 예측된 양안 불일치 맵을 직접 생성한다. 소스 이미지가 양안 스테레오 카메라의 보정된 쌍에 의해 캡처된 한쌍의 스테레오 이미지 중 하나인 경우, 생성된 양안 불일치 값은 캡처된 소스 이미지 내의 검출된 오브젝트 또는 피처의 이미지 위치의 차이 및 대응하는 개념적인 양안 스테레오 뷰에서의 오브젝트 또는 피처의 예측된 이미지 위치를 나타낸다. 깊이 데이터 생성 모듈(5)은 불일치 예측기(9)에 의해 출력된 양안 불일치 맵으로부터 깊이 정보를 산출한다.
- [0025] CNN(11)은 처리 노드들의 동적 구조화된 배열을 포함하며, 각 노드는 대응하는 가중치 파라미터를 가진다. CNN(11)을 정의하는 구조 및 가중치는 훈련 스테이지 동안 훈련 모듈(13)에 의해 업데이트된다. 본 실시예에서, CNN(11)의 처리 노드들은 3개의 주요 구성들로 배열된다:
- [0026] - 노드 및 레이어를 포함하여 입력 이미지 데이터를 처리하며 입력 이미지 내의 오브젝트 또는 피처를 나타내는 인코딩된 데이터를 출력하는 인코더(12);
- [0027] - 노드 및 레이어를 포함하여 인코더(12)로부터 인코딩된 데이터를 처리하고, 증가된 공간 해상도로 스케일링된 데이터를 출력하기 위해 업 컨볼루션 및 업 샘플링을 수행하며, 불일치 예측기(9)의 의해 출력된 입력 인코딩된 데이터로부터 예측된 불일치 맵과 같은, 예측된 불일치 맵을 출력하고, 예측된 불일치 맵을 입력 이미지 데이터에 적용하여 투영된 뷰를 출력하는 디코더(14); 및
- [0028] - 노드 및 레이어를 포함하여 CNN(11)을 업데이트하는 데 사용되는 훈련 손실을 계산하며, 상기 훈련 손실은 디코더(14)에 의해 출력된 불일치 맵으로부터 계산된 불일치 평탄도와 좌-우 불일치 일관성 비용 항목 및 투영된 뷰와 대응하는 입력 뷰의 비교로부터 계산된 외형 매칭 비용 항목을 포함하는 손실 모듈(19).
- [0029] 하기에서 더 상세히 설명되는 바와 같이, 훈련 모듈(13)은 예를 들어 트레이닝 이미지의 데이터베이스(17)로부터 검색된 양안 스테레오 이미지 쌍(15)에 기초하여 컨볼루션 뉴럴 네트워크(CNN) 모듈(11)을 훈련한다. 양안 스테레오 이미지 쌍(15)은 공지된 카메라 초점 거리 및 공지된 베이스라인 거리를 두고 각각의 양안 스테레오 카메라에 의해 동시에 캡처된 좌측 뷰(15a) 및 우측 뷰(15b)를 포함하고, 양안 스테레오 이미지 쌍에 의하여 깊이 데이터가 불일치 예측기(9)에서 출력된 예측된 양안 불일치 값으로부터 계산될 수 있다. 훈련 모듈(13)은 CNN 모듈(11)의 손실 모듈(19)에 의해 구현된 손실 함수를 최적화하고, 그 결과 단일 소스 이미지의 컬러 픽셀 값들로부터 직접 예측된 양안 불일치 맵을 정확하고 효율적으로 생성하도록 불일치 예측기(9)를 훈련한다.
- [0030] CNN 모듈(11), 트레이닝 모듈(13) 및 깊이 데이터 생성기 모듈(5)은 단일 모듈로 결합되거나 추가적인 모듈로 분할될 수 있고, 이미지 처리 모듈(3)은 훈련된 CNN 모듈(11)의 모델 데이터를 저장하기 위한 메모리(21)를 포함한다. 시스템(1)은 설명의 명료성을 위해 도 1에 도시되지 않은 컴퓨팅 시스템/장치에 일반적으로 존재하는 다른 컴포넌트, 서브 컴포넌트, 모듈 및 디바이스 또한 포함할 수 있다.

[0031] 이미지 처리 시스템(3)에 의해 출력된 깊이 정보는 추가 데이터 처리를 위해 하나 이상의 깊이 데이터 처리 모듈(23)에 제공될 수 있다. 깊이 데이터 처리 모듈(23)은 처리된 깊이 데이터에 기초하여 출력 장치(25)에 데이터 및/또는 제어 신호를 출력하도록 구성될 수 있다. 깊이 데이터 처리 모듈의 특성과 배열은 시스템(1)의 구현 환경에 따라 달라질 것이다. 순전히 예시적인 구체적 구현 방법으로: 깊이 맵을 컴퓨터 그래픽에서 합성 오브젝트 삽입과 관련하여 캡처된 이미지 데이터로부터 예측; 전산 사진에서 합성 피사체 심도를 결정; 로봇 파지를 위한 제어 명령을 생성; 인체 자세 추정에서 단서로서의 깊이를 출력; 인간 컴퓨터 상호 작용에서 손 포즈 추정에 대한 강한 단서를 결정; 영화 비디오 데이터를 위한 2D / 3D 자동 변환; 자율주행 차량용 저비용 장애물 회피 센서; 소형 폼 팩터, 단일 카메라, 깊이 감지, 수술용 내시경; 단일 카메라 3D 재구성; VR 헤드셋에 대한 향상된 포즈 추정; 장애물 회피 및 경로지도 작성; 물체 계측을 위한 크기 및 부피 추정하는 것이 있다. 훈련 데이터(17)는 특정 구현 환경에 세부적인 뷰들의 스테레오 이미지 쌍(15)을 포함할 수 있다는 것을 이해하여야 한다.

[0032] 도 2는 본 실시예에 따른 예시적인 CNN의 디코더(14) 및 훈련 손실 모듈(19) 섹션의 개략도이다. CNN(11)의 예시적인 레이어들은 아래의 표 1에 설명되어 있으며, 이는 엔. 마이어, 이. 일그, 피. 하우스저, 피. 피셔, 이. 그래머즈, 에이. 도스비츠키 및 티. 브록스, "불일치, 옵티컬 플로우 및 장면 플로우 추정을 위한 컨볼루션 네트워크 학습을 위한 대규모 데이터 세트", CVPR 2016. (N. Mayer, E. Ilg, P. Hausser, P. Fischer, D. Cremers, A. Dosovitskiy, and T. Brox, "A Large Dataset To Train Convolutional Networks For Disparity, Optical Flow, And Scene Flow Estimation", CVPR 2016.)에 의한 완전한 컨볼루션 아키텍처(fully convolutional architecture)에 기초하나, 실제 측정 깊이 데이터(ground truth depth data)를 요구하지 않고 네트워크를 훈련시킬 수 있는 몇 가지 수정을 포함하도록 조정되었다. 도시된 예시에서, CNN은 훈련 단계에서 시스템에 의해 학습된 3,100 만 개의 매개 변수로 구성되며, 여기서 'k'는 커널 크기, 's'는 스트라이드, 'channels'는 각 레이어의 입력 및 출력 채널 수, 'in' 및 'out'은 입력 이미지에 대한 각 레이어의 입력 및 출력 다운스케일 계수, 및 'input'은 '+'가 연결을 의미하고 '\*'는 해당 레이어의 2x 업 샘플링에 해당하는 각 레이어의 입력에 해당한다.

[0033] [표 1]

Encoder layer	k	s	channels	in	out	Input layer
conv1	7	2	3/32	1	2	left
conv1b	7	1	32/32	2	2	conv1
conv2	5	2	32/64	2	4	conv1b
conv2b	5	1	64/64	4	4	conv2
conv3	3	2	64/128	4	8	conv2b
conv3b	3	1	128/128	8	8	conv3
conv4	3	2	128/256	8	16	conv3b
conv4b	3	1	256/256	16	16	conv4
conv5	3	2	256/512	16	32	conv4b
conv5b	3	1	512/512	32	32	conv5
conv6	3	2	512/512	32	64	conv5b
conv6b	3	1	512/512	64	64	conv6
conv7	3	2	512/512	64	128	conv6b
conv7b	3	1	512/512	128	128	conv7

[0034]

Decoder layer	k	s	channels	in	out	Input layer
upconv7	3	2	512/512	128	64	conv7b
iconv7	3	1	1024/512	64	64	upconv7+conv6
upconv6	3	2	512/512	64	32	iconv7
iconv6	3	1	1024/512	32	32	upconv6+conv5
upconv5	3	2	512/256	32	16	iconv6
iconv5	3	1	512/256	16	16	upconv5+conv4
upconv4	3	2	256/128	16	8	iconv5
iconv4	3	1	128/128	8	8	upconv4+conv3
disp4	3	1	128/2	8	8	iconv4
upconv3	3	2	128/64	8	4	iconv4
iconv3	3	1	66/64	4	4	upconv3+conv2+disp4*
disp3	3	1	64/2	4	4	iconv3
upconv2	3	2	64/32	4	2	iconv3
iconv2	3	1	34/32	2	2	upconv2+conv1+disp3*
disp2	3	1	32/2	2	2	iconv2
upconv1	3	2	32/16	2	1	iconv2
iconv1	3	1	18/16	1	1	upconv1+disp2*
disp1	3	1	16/2	1	1	iconv1

[0035]

[0036]

도시된 바와 같이, CNN(11)은 인코더(12) (conv1 내지 conv7b를 포함하는) 및 디코더(14) (upconv7 내지 disp1의 레이어를 포함하는)를 포함한다. 디코더(14)는 더 높은 해상도의 세부 사항을 해결하기 위해, 당 업계에 공지된 바와 같이, 인코더의 활성화 블록으로부터 스킵 연결을 구현할 수 있다. 도 2에서, C는 Convolution 연결, UC는 Up-Convolution 연결, S는 Bi-linear Sampling 연결, US는 Up-Sampling 연결을 나타낸다. 본 예시적인 실시예에서, 불일치 예측들은 4개의 상이한 스케일(disp4 내지 disp1로 라벨링 됨)에서 출력되고, 이는 후속 스케일들 각각에서 공간 해상도를 증가시킨다. 네트워크를 훈련할 때, 출력 스케일들 각각에서, 각각의 입력 이미지 뷰(예를 들어, 좌측 및 우측 뷰)에 대해 2개의 불일치 맵이 s로 표시된 것처럼 예측된다. 하나의 불일치 맵은 레이어에 대한 입력(예를 들어, 좌측 뷰의 인코딩된 데이터에 정렬된 좌측-우측 불일치 맵,  $d^l$ )과 정렬된다. 다른 불일치 맵은 대응하는 스테레오 파트너(예를 들어, 대응하는 투영된 우측 뷰에 정렬되는 투영된 우측-좌측 불일치 맵,  $d^r(d^l)$ )과 정렬된다. 디코더(14) 및 손실 모듈(19)에 의한 프로세싱은 4개의 상이한 출력 스케일들 각각에서 반복된다.

[0037]

중요한 이점은 훈련된 시스템(3)이 양안 카메라로부터의 불일치를 예측하고 서로 일관성을 유지하도록 강화함으로써 우수한 깊이 맵을 생성한다는 것이다. 좌측 뷰의 픽셀을 사용하여 우측 뷰를 생성하면 우측 뷰와 정렬된 불일치 맵이 생성된다(반대의 경우도 마찬가지이다). 훈련 모듈(13)은 예측된 불일치 맵의 정렬 정렬을 소스 입력 이미지(이 실시예에서, 좌측 뷰(15a))에 최적화하는 것을 목표로 한다. 훈련 동안, 훈련 모듈(13)은 좌측 및 우측 스테레오 이미지(15a, 15b) 모두에 액세스하고, 훈련 모듈(13)은 CNN(11)을 훈련하여 좌측-우측 및 우측-좌측 불일치 맵을 추정할 뿐만 아니라, 각각의 추정된 불일치 맵으로부터 대응하는 투영된 우측-좌측 및 좌측-우측 불일치 맵을 결정하고, 이들 사이의 일관성을 강화한다. 훈련 모듈(13)의 추가적인 최적화 목표는 이미지 재구성 에러를 최소화하기 위해 픽셀을 쉬프트할 수 있는 불일치 맵을 학습함으로써 대응하는 좌측 및 우측 뷰를 재구성하도록 CNN(11)을 훈련시키는 것이다. 이러한 방식으로, 교정된 양안 카메라의 쌍으로부터 훈련 이미지들이 주어지면, 이미지 처리 시스템(3)은 다른 뷰가 주어진 이미지를 재구성할 수 있는 기능을 학습하고, 그렇게 함으로써, 이미징되고 있는 장면의 형태의 예측 또는 추정을 가능하게 하는 훈련 모델(즉, CNN(11))을 생성한다. 단일 훈련 이미지(I)(예를 들어, 훈련 스테레오 이미지 쌍(15)의 좌측 뷰(15a))가 주어지면, 이미지 처리 시스템(3)은 또한 훈련 동안 이미지 재구성 문제로서 깊이 추정을 처리하는 픽셀 당 장면 깊이를 예측할 수 있는 함수,  $\hat{d} = f(I)$ ,를 학습한다.

[0038]

일 실시예에 따른 이미지 처리 시스템(3)의 일부를 형성하는 구성 요소에 대한 개략적인 설명을 상술하였다. 이들 구성 요소들의 동작, CNN(11)이 예를 들어 해당 실제 측정 깊이 정보(ground truth depth information)의 형태로 지도를 요구하지 않고, 단지 스테레오 이미지 쌍들에 대해 훈련되는 것을 가능하게 하는 일 실시예에 따

른 단일 이미지 깊이 예측 CNN(11)을 훈련하는 과정에 대한 상세한 설명을 이제 도 3의 순서도를 참조하여 서술하도록 한다. 이 흐름도의 다양한 단계들이 순차적으로 제공되고 설명되지만, 단계들의 일부 또는 전부가 상이한 순서로 실행될 수 있고, 결합되거나 생략될 수 있으며, 단계들의 일부 또는 전부가 병행하여 실행될 수 있음을 이해해야 한다. 또한, 하나 이상의 예시적인 실시예에서, 후술되는 하나 이상의 단계는 생략되거나, 반복되거나, 및/또는 다른 순서로 수행될 수 있다.

[0039] 본 발명의 실시예에 따른, CNN(11)의 처리 노드 및 레이어의 예시적인 구조화된 배열을 개략적으로 도시한 블록 순서도를 도시하는 도 4가 또한 참조된다. 간략화를 위해, 도 2에 도시된 업 컨볼루션(UC) 및 업 샘플링(US) 레이어는 도 4에서 생략되었지만, UC 및 US 레이어로부터 스케일링된 출력은 각각의 예측된 불일치 및 각각의 계산된 비용 요소에 대해 아래 첨자  $s$ 로 표시된다는 것을 이해해야 한다.

[0040] 도 3에 도시된 바와 같이, 한 쌍의 훈련 이미지(15)에 대한 트레이닝 프로세스의 반복은 CNN(11)이 입력 스테레오 쌍 중 하나의 뷰(이 실시 예에서는 좌측 뷰)의 컬러 이미지 데이터를 수신하는 단계(S3-1(L))에서 시작한다. 이 실시예에서, CNN(11)은 단계(S3-1(R))에서 우측 뷰의 컬러 이미지 데이터도 수신한다. 훈련 모듈(13)은 메모리(17)에 저장된 훈련 데이터로부터 같은 시간에 캡처되고, 보정된 스테레오 쌍의 좌측 및 우측 컬러이미지에 대응하는 2개의 이미지( $I^l$  및  $I^r$ )를 검색하고, 이미지 데이터를 CNN(11)의 하나 이상의 입력 노드(미도시)에 전달한다. CNN(11)은 병렬로 다수의 훈련 이미지 쌍을 유리하게 수신하고 처리하도록 구성될 수 있음을 이해하여야 한다. 필수적이지는 않지만, 바람직하게, 이미지의 스테레오 쌍(15)은 정류되어, 본 기술 분야에서 알려진 바와 같이 정의된 변환 프로세스를 사용하여 이미지가 공통 이미지 평면 상에 투영된다.

[0041] 단계(S3-3)에서, 좌측 뷰의 입력 이미지 데이터는 인코더(12)의 컨볼루션 레이어를 통과하여 인코딩된 입력 데이터, 예를 들어 입력 이미지의 식별된 오브젝트 또는 피처의 복잡한 피처 벡터를 나타내는 입력 데이터를 생성한다. 좌측 뷰 이미지(15a)로부터 직접 깊이를 예측하려고 하는 대신에, CNN(11)은 대응 필드를 찾도록 훈련된다. 대응 필드는 본 실시예에서 좌측 뷰 이미지(15a)에 적용될 때 CNN(11)의 우측 뷰 프로젝터(415a)가 투영된 우측 뷰 이미지를 재구성할 수 있게 하는(또는, 그 반대로) 예측된 좌측-우측 불일치 맵( $d^l$ )이다. 따라서, 단계(S3-5)에서, 단계(S3-3)에서 출력된 인코딩된 데이터는 현재 구조 및 가중치에 기초하여 예측된 좌측-우측 불일치 맵( $d^l_s$ )의 데이터 값을 출력하는 좌측 뷰 불일치 예측기(307a)의 프로세싱 노드를 통과한다. 이하에서 설명하는 바와 같이, CNN(11)은 입력 양안 스테레오 쌍의 각각의 이미지에 대해, 대응하는 불일치 값을 예측함으로써 불일치 맵을 입력 데이터로부터 예측하도록 훈련된다. 스테레오 쌍 내의 각각의 이미지에 대한 예측된 불일치 값 사이의 일관성을 강화하는 비용 함수에 기초하여 CNN(11)은 업데이트된다. 따라서, 단계(S3-5)에서, 인코딩된 데이터는 또한 현재 구조 및 가중치에 기초하여 예측된 우측-좌측 불일치 맵( $d^r_s$ )의 데이터 값을 출력하는 좌측 뷰 불일치 예측기(307b)의 프로세싱 노드를 통과한다.

[0042] 선택적으로, 단계(S3-7)에서, 손실 모듈(13)의 L->R 불일치 평활화 노드(413a)에 의해 좌측-우측 불일치 평활도 비용( $(C^r_{ds})_s$ )이 예측된 좌측-우측 불일치 맵( $d^r$ )으로부터 계산될 수 있다. 마찬가지로, 단계(S3-7)에서, 손실 모듈(13)의 R->L 불일치 평활화 노드(413b)에 의해 우측-좌측 불일치 평활도 비용( $(C^l_{ds})_s$ )이 예측된 우측-좌측 불일치 맵( $d^l$ )으로부터 계산될 수 있다. 훈련 손실 함수의 계산된 평활도 비용 요소는 각각의 예측된 불일치 맵이 불일치 그래디언트( $\partial d$ )에 대한 L1 페널티를 사용하여 국부적으로 평활화 되는 것을 장려한다. 예를 들어, 예측된 좌측 불일치 맵( $d^l$ )으로부터 계산된 평활도 비용은 하기 수학식 1과 같이 공식화될 수 있다.

[0043] [수학식 1]

$$C^l_{ds} = \frac{1}{N} \sum_{i,j} |\partial_x d^l_{ij}| e^{-\eta \|\partial_x I^l_{ij}\|} + |\partial_y d^l_{ij}| e^{-\eta \|\partial_y I^l_{ij}\|}$$

[0044] 여기서,  $\eta$  은 1.0으로 설정되고, 깊이 불연속성이 종종 이미지 그래디언트에 발생하므로, 평활도 비용은 대응하는 이미지 그래디언트( $\partial I$ )를 사용하는 예지 인식 용어로 가중될 수 있다.

[0046] 단계(S3-9(L))에서, R->L 불일치 프로젝터(409a)는 예측된 좌측-우측 불일치 맵( $d^r_s$ )의 데이터 값을 샘플링하고,



예측된 우측-좌측 불일치 맵( $d^l_s$ )에 샘플링된 데이터를 적용하여 투영된 우측-좌측 불일치 맵( $d^l(d^r)_s$ )을 생성한다. 명확하게 하기 위해, 예측된 좌측-우측 불일치 값의 처리는 (L)로 표시된 단계를 참조하여 설명될 것이다. (R)로 표시된 바와 같이, 대응하는 번호가 부여된 처리 단계는 우측-좌측 불일치 값에 대해 반영한다. 이 실시예에서, 불일치 프로젝터(409)는 불일치 맵을 사용하여 입력 데이터를 샘플링하기 위해 예를 들어, 엠. 자데르버그, 케이. 시몬얀, 에이. 지세르만, 케이. 카부큐오글루, "공간 트랜스포머 네트워크", NIPS 2015. (M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial Transformer Networks", NIPS 2015.)에서 공지된 바와 같은, 공간 트랜스포머 네트워크(STN)로부터의 이미지 샘플러에 기초한 이미지 샘플링 기능을 구현한다. 공간 트랜스포머 네트워크(STN)는 출력 픽셀이 4개의 입력 픽셀의 가중 합인 이중선형(bilinear) 샘플링을 사용할 수 있다. 시에 외.(Xie et al.) 및 가그 외.(Garg et al.)의 전술한 접근법과는 대조적으로, 본 실시예에서 사용된 이중선형(bilinear) 샘플러는 국부적으로 완전 미분 가능(locally fully differentiable)하고 CNN(11)의 완전한 컨볼루션 아키텍처에 완전히 통합(integrates seamlessly)될 수 있다. 이것은 CNN(11)이 최적화 비용 함수의 단순화 또는 근사화를 필요로 하지 않는다는 것을 의미한다.

[0047] 보다 견고한 결과를 생성하기 위해, CNN(11)은 네트워크의 컨볼루션 손실 모듈(13) 부분에 대한 입력으로 좌측 뷰 이미지 데이터(15a)에만 기초하여, 좌측 및 우측 이미지 불일치를 예측하도록 훈련된다. 이에 따라, 단계(S3-9(L))에서, 단계(S3-5)에서 좌측 뷰 불일치 예측 노드(407a)가 출력한 예측 좌측 불일치 맵( $d^l$ )에 기초하여, CNN(11)의 투영된 우측 불일치 예측 노드(409a)는 투영된 우측 불일치 맵( $d^l(d^r)$ )을 출력한다. 일관성을 보장하기 위해, 손실 모듈(13)은 모델(11)의 일부로서 L1 좌측-우측 불일치 일관성 패널티를 포함한다. 이러한 비용은 예측된 좌측 뷰 불일치 맵( $d^l$ )을 투영된 우측 뷰 불일치 맵( $d^r(d^l)$ )과 동일하게 유도하기 위해 제공된다. 따라서, 단계(S3-11(L))에서, L-R 불일치 일관성 손실 노드(411a)는 좌측 일관성 비용을 하기 수학적식 2와 같이 계산한다:

[0048] [수학적식 2]

$$C_{lr}^l = \frac{1}{N} \sum_{i,j} |d_{ij}^l - d_{ij+d_{ij}^l}^r|$$

[0049]

[0050] 단계(S3-13(L))에서, CNN(11)의 조밀 스케일러(coarse-to-fine scaler, 405a)는 스케일( $s_1$  내지  $s_n$ )에서, 좌측 뷰의 스케일링된 이미지 데이터를 생성하여 출력한다. 본 발명의 예시적인 실시예에서  $n=4$ 일 수 있다. 각각의 스케일(s)에 대해, 좌측 뷰의 대응하는 스케일링된 이미지 데이터( $I_s^l$ )는 프로세싱을 위해 디코더(14)의 우측 뷰 프로젝터(415a)로 전달된다. 단계(S3-15(L))에서, 우측 뷰 프로젝터(415a)는 스케일링된 좌측 뷰 이미지 데이터( $I_s^l$ )로부터 픽셀을 샘플링하여 투영된 인접한 스테레오 이미지를 생성한다. 이 실시예에서, 뷰 프로젝터(415) 또한 입력 불일치 맵을 사용하여 입력 데이터를 샘플링하기 위해, 상술한 바와 같은 공간 트랜스포머 네트워크(STN)로부터 이미지 샘플러를 구현한다.

[0051] 단계(S3-15(L))에서, 선택적으로, CNN(11)의 우측 뷰 프로젝터(415a)는 예측된 좌측-우측 불일치( $d^r$ )를 입력된 스케일링된 좌측 뷰 이미지 데이터( $I_s^l$ )에 적용함으로써 투영된 우측 뷰 이미지를 재구성할 수 있다. 이러한 과정은 하기 수학적식 3과 같이 공식화될 수 있다.

[0052] [수학적식 3]

$$\arg \min_{a_r} \|I^r - I^l(d^r)\|$$

[0053]

[0054] 여기서  $d$ 는 모델(11)이 예측하도록 훈련된 픽셀 당 스칼라 값인 이미지 불일치에 대응한다. 재구성된 이미지( $I^r(d^r)$ )은 간략화를 위해  $\hat{I}^r$ 으로 나타낸다. 단계(S3-13(R)) 및 S3-15(R))에서, 투영된 좌측 뷰 이미지는 예측된 우측-좌측 불일치 맵( $d^l$ )을 입력 스케일링된 우측 뷰 이미지 데이터( $I_s^l$ )에 적용하여 유사하게 생성될 수 있다.

[0055] 단계(S3-17(L))에서, 외형 매칭 비용은 R 외형 매칭 손실 노드(417a)에서 계산될 수 있다. 외형 매칭 비용은 L1 및 단일 스케일 구조적 유사성, SSIM, 광도계의 용어로서, 입력 이미지( $I_{ij}^l$ )와 이의 재구성( $\hat{I}_{ij}^l$ ) 사이의 이미지

재구성 비용의 조합으로 하기 수학식 4와 같이 계산될 수 있다.

[수학식 4]

$$C_{ap}^l = \frac{1}{N} \sum_{i,j} \alpha \frac{1 - \text{SSIM}(I_{ij}^l, \tilde{I}_{ij}^l)}{2} + (1 - \alpha) \|I_{ij}^l - \tilde{I}_{ij}^l\|$$

여기서 N은 이미지의 픽셀 수이다. 예시적인 실시예에서, 가우시안 대신에 3 X 3 블록 필터를 갖는 단순화된 SSIM이 사용되고,  $\alpha$ 는 0.85로 설정된다. SSIM 용어의 계산은 예를 들어, 지.왕, 에이. 씨. 보빅, 에이치. 알. 세이키, 및 이. 피. 시몬첼리, "이미지 품질 평가: 오류 가시성에서 구조적 유사성까지", 이미지 처리 트랜잭션 2004. (Z.Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility To Structural Similarity", Transactions on Image Processing 2004.)와 같이 당 업계에 공지되어 있으며, 이에 대한 상세한 설명은 생략하도록 한다. 단계(S3-17(R))에서, 좌측 외형 매칭 비용( $C_{ap}^l$ )은 좌측 뷰 프로젝터(415b)에 의해 출력된 투영된 좌측 뷰 및 스케일러(405)에서 출력된 대응하는 스케일링된 좌측 뷰 이미지로부터 유사하게 계산될 수 있다.

단계(S3-19)에서, 통해 좌측 뷰 컬러 영상(15a)과 우측 뷰 컬러 영상(15b)이 CNN(11)을 통과하면, 손실 모듈(13)의 트레이닝 손실 노드(419)는 현재 스케일에서의 스테레오 쌍의 훈련에 대한 훈련 손실을 계산한다. 본 실시예에서, 스케일링된 훈련 손실은 단계(S3-7(L) 및 (R))에서 출력된 불일치 평활도 비용, 단계(S3-11(L) 및 (R))에서 출력된 불일치 일관성 비용, 및 단계(S3-17(L) 및 (R))에서 출력된 외형 매칭 비용의 가중치 조합으로 계산된다. 계산된 3가지 비용 항목의 가중치 조합은 하기와 같은 수학식 5와 같이 공식화될 수 있다.

[수학식 5]

$$C_s = \alpha_{ap}(C_{ap}^l + C_{ap}^r) + \alpha_{ds}(C_{ds}^l + C_{ds}^r) + \alpha_{lr}(C_{lr}^l + C_{lr}^r)$$

여기서,  $C_{ap}$ 는 재구성된 이미지가 해당 학습 입력과 유사하도록 조장하고,  $C_{ds}$ 는 원활한 불일치를 강화하며,  $C_{lr}$ 은 왼쪽 및 오른쪽 이미지의 예측된 불일치를 일관되게 유지하려고 시도한다. 각각의 주요 항목은 왼쪽 및 오른쪽 이미지 변형을 모두 포함한다. 3가지 훈련 비용 요소를 모두 포함하는 이러한 예시적인 실시예에서, 좌측 뷰 이미지(15a)는 CNN(11)을 항상 통과한다. 훈련 모듈(13)은 훈련 과정에서 대응하는 우측 뷰 이미지(15b)에 액세스하기 때문에, CNN(11) 또한 이의 기준 프레임에서 불일치 맵을 예측할 수 있다. 외형 매칭 비용 요소가 구현되지 않을 때 오른쪽 뷰 이미지 데이터가 CNN(11)을 통과할 필요는 없다는 것을 이해해야 한다.

단계(S3-21)에서, CNN(11)의 디코더(14)는 현재의 스케일에서 인코더(12)에 의해 출력된 데이터의 업-컨볼루션을 수행할 뿐만 아니라, 불일치 예측기(407)에 의해 출력된 예측된 불일치 맵의 업-샘플링도, 단계(S3-3)에서 설명한 바와 같이 다음 스케일에 대한 스케일링된 훈련 손실을 계산하기 위한 후속하는 구조화된 처리 노드 세트의 입력으로, 수행한다. 미리 정의된 각각의 스케일에 대한 스케일링된 훈련 손실의 계산 이후, 최종 손실 총량은 단계(S3-23)에서 손실 모듈(13)의 합산 노드(421)에 의해 개별 스케일링된 손실( $C_s$ )의 가중치 합으로 하기 수학식 6과 같이 계산된다.

[수학식 6]

$$C = \sum_{s=1}^4 \lambda_s C_s$$

여기서,  $\lambda_s$ 은 훈련 모듈(13)이 훈련 동안 상이한 출력 스케일의 상대적 중요도를 가중하도록 구성되게 한다.

예시적인 실시예에서, 상이한 손실 컴포넌트의 가중치는  $\alpha_{ap}=1$  및  $\alpha_{lr}=1$ 로 설정된다. 가능한 출력 불일치는 스케일링된 시그모이드 비선형성을 사용하여 0과  $d_{\max}$  사이로 제한되며, 여기서,  $d_{\max}$  = 주어진 출력 스케일에서 이미지 폭의 0.3 배이다. 멀티 스케일 출력의 결과로서, 인접한 픽셀들의 전형적인 불일치는 (CNN(11)이 출력을 2가지 팩터로 업 샘플링함에 따라) 각 스케일 사이에서 2가지 팩터에 의해 상이할 것이다. 이것을 보정하기 위해, 트레이닝 모듈(13)은 각각의 스케일에 대해 불일치 평활도 항목( $\alpha_{ds}$ )을  $r$ 로 스케일링 하여, 각 레벨에서 동일한 평활화를 획득한다. 따라서,  $\alpha_{ds} = 0.1/r$  이고,  $r$ 은 CNN(11)에 전달되는 입력 이미지의 해상도에 대한

대응하는 레이어의 다운 스케일링 팩터이다(표 1에서).

[0068] 단계(S3-25)에서, 훈련 모듈(13)은 단계(S3-21)에서 합산 노드(421)에 의해 계산된 최종 훈련 손실 총량의 가중된 컴포넌트를 역-전파시킴으로써 CNN(11)을 업데이트한다. CNN(11)의 비선형성에 대해, 이 기술 분야에 공지되어 있는 바와 같이, 일반적으로 사용되는 ReLU(rectified liner units) 대신 지수 선형 유닛이 사용될 수 있다. 예시적인 실시예에서, CNN(11)은 다. 킹마 및 제이. 바, "아담: 확률적 최적화를 위한 방법", arXiv preprint, arXiv:1412.6980, 2014.(D. Kingma and J. Ba, "Adam: A method for stochastic optimization", arXiv preprint, arXiv:1412.6980, 2014.)에 설명된 기술에 기초하여, 핸디캡이 없는 동일선산의 출발선(scratch)부터 50 에포크(epochs)까지 훈련될 수 있으며, 여기서  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ , 및  $\epsilon = 10^{-8}$ 이다.  $\lambda = 10^{-4}$ 의 초기 학습 속도가 사용되며, 속도는 첫번째 30개의 에포크에 대해 일정하게 유지되어 끝날 때까지 10개 에포크마다 반으로 줄어든다. 훈련 모듈(13)은 순차적인 업데이트 스케줄을 사용하여 CNN(11)을 업데이트하도록 구성될 수 있으며, 저해상도 이미지 스케일이 먼저 최적화된다. 그러나, 본 발명의 발명자들은 4개의 스케일 모두를 동시에 최적화하는 것이 더욱 안정적인 컨버전스를 유도한다는 것을 알게 되었다. 유사하게, 상이한 가중치가 불안정한 컨버전스를 야기하는 경우, 각 스케일 손실의 동일한 가중치가 사용될 수 있다.

[0069] 도 6은 일 실시예에 따라 훈련된 CNN(11)을 사용하여 단일 소스 이미지로부터 깊이 데이터를 생성하고 처리하는 예시적인 과정에 대한 순서도이다. 단계(S6-1)에서, 단일 소스 이미지의 컬러 이미지 데이터는 예를 들어 카메라(7)로부터 깊이 데이터 생성기(5)에 의해 수신된다. 단계(S6-3)에서, 훈련된 CNN(11)을 통한 단일 순방향 패스를 사용하여, 좌측 이미지( $d^l$ )에 대한 가장 큰 스케일의 불일치는 훈련된 L-R 뷰 불일치 예측기(407a)에 의해 예측된 불일치 맵으로서 출력된다(표 1의 disp1에 해당). CNN(11)에서의 조밀 스케일러(405)에 의한 업 샘플링의 결과로서, 출력된 예측된 불일치 맵은 입력 이미지와 동일한 해상도이다. 좌-우 불일치( $d^l$ )는 깊이 데이터 생성 단계에서 사용되지 않는다는 것을 알아야 한다.

[0070] 단계(S6-5)에서, 깊이 데이터 생성기(5)는 단계(S6-3)에서 출력된 예측된 불일치 맵으로부터 계산되는, 소스 이미지의 각 픽셀에 대한 예측 깊이 값으로 구성된 깊이 이미지를 생성한다. 훈련 데이터(15)를 캡처하는 데 사용된 스테레오 카메라들 사이의 기본 거리(b) 및 관련 카메라 초점 거리(f)가 주어지면, 깊이 데이터 생성기(5)는 하기와 같은 수학적 7과 같이 예측된 불일치로부터 추정된 깊이 값을 복구할 수 있다:

[0071] [수학식 7]

$$\hat{d} = b \frac{f}{a}$$

[0072]

[0073] 단계(S6-7)에서, 깊이 이미지는 깊이 데이터 처리 모듈(23)로 전달되어 시스템(1)의 특정 구현 환경에 따라 처리된다.

## [0074] 컴퓨터 시스템

[0075] 이미지 처리 시스템(3) 및/또는 이미지 처리 시스템(3)의 개별 모듈과 같은 여기에 기술된 엔티티는 도 6에 도시된 바와 같은 컴퓨터 시스템(1000)과 같은 컴퓨터 시스템에 의해 구현될 수 있다. 본 발명의 실시예는 이러한 컴퓨터 시스템(1000)에 의한 실행을 위한 프로그램 가능 코드로서 구현될 수 있다. 이 설명을 읽은 후에, 다른 컴퓨터 시스템 및/또는 컴퓨터 아키텍처를 사용하여 본 발명을 구현하는 방법은 당업자에게 명백해질 것이다.

[0076] 퍼스널 컴퓨터, 랩탑, 컴퓨팅 단말기, 스마트 폰, 태블릿 컴퓨터 등일 수 있는 컴퓨터 시스템(1000)은 프로세서(1004)와 같은 하나 이상의 프로세서를 포함한다. 프로세서(1004)는 특수 목적 또는 범용 디지털 신호 프로세서를 포함하지만 이에 한정되지 않는 임의의 유형의 프로세서일 수 있다. 프로세서(1004)는 통신 인프라 구조(1006)(예를 들어, 버스 또는 네트워크)에 접속된다. 다양한 소프트웨어 구현이 이러한 예시적인 컴퓨터 시스템에 의해 설명된다. 이 설명을 읽은 후에, 다른 컴퓨터 시스템 및/또는 컴퓨터 아키텍처를 사용하여 본 발명을 구현하는 방법은 당업자에게 명백해질 것이다.

[0077] 컴퓨터 시스템(1000)은 또한 하나 이상의 입력 장치(들)(1005)에 연결된 사용자 입력 인터페이스(1003) 및 하나 이상의 디스플레이 (들)(1009)에 연결된 디스플레이 인터페이스(1007)를 포함한다. 입력 장치들(1005)은 예를 들어, 마우스 또는 터치 패드와 같은 포인팅 장치, 키보드, 저항성 또는 용량 성 터치 스크린과 같은 터치 스크린 등을 포함할 수 있다. 이 설명을 읽은 후에, 예를 들어 통합된 입력 및 디스플레이 컴포넌트를 갖는 착탈식 전자 디바이스를 사용하여, 다른 컴퓨터 시스템 및/또는 컴퓨터 아키텍처를 사용하여 본 발명을 구현하는 방법



은 당업자에게 명백해질 것이다.

- [0078] 또한, 컴퓨터 시스템(1000)은 주 메모리(1008), 바람직하게는 랜덤 액세스 메모리 (RAM)를 포함하고, 또한 2 차 메모리(1010)를 포함할 수 있다. 제 2 메모리(1010)는 예를 들어, 플로피 디스크 드라이브, 자기 테이프 드라이브, 광 디스크 드라이브 등을 나타내는 하드 디스크 드라이브(1012) 및 / 또는 착탈식 저장 드라이브(1014)를 포함할 수 있다. 착탈식 저장 드라이브(1014)는 잘 알려진 방식으로 착탈식 저장 유닛(1018)로부터 판독하고 및 / 또는 착탈식 저장 유닛(1018)에 기록한다. 착탈식 저장 유닛(1018)은 플로피 디스크, 자기 테이프, 광 디스크 등을 나타내며, 착탈식 저장 드라이브(1014)에 의해 판독되어 기록된다. 인식 할수 있는 바와 같이, 착탈 가능한 저장 유닛(1018)은 컴퓨터 소프트웨어 및/또는 데이터가 저장된 컴퓨터 사용 가능한 저장 매체를 포함한다.
- [0079] 다른 구현 예에서, 2차 메모리(1010)는 컴퓨터 프로그램 또는 다른 명령들이 컴퓨터 시스템(1000)에 로딩될 수 있게 하는 다른 유사한 수단을 포함할 수 있다. 이러한 수단은 예를 들어, 착탈식 저장 유닛(1022) 및 인터페이스(1020)를 포함할 수 있다. 이러한 수단의 예는 프로그램 카트리지 및 카트리지 인터페이스 (이전에 비디오 게임 장치에서 발견된 것과 같은), 착탈식 메모리 칩 (EPROM 또는 PROM 또는 플래시 메모리와 같은) 및 관련 소켓, 및 다른 착탈식 저장 유닛(1022) 및 인터페이스(1020)를 포함하며, 이는 소프트웨어 및 데이터가 착탈 가능 저장 유닛(1022)으로부터 컴퓨터 시스템(1000)으로 전송 되도록 한다. 대안적으로, 컴퓨터 시스템(1000)의 프로세서(1004)를 사용하여 프로그램이 실행될 수 있고 및/또는 데이터가 착탈식 저장 유닛(1022)으로부터 액세스 될 수 있다.
- [0080] 컴퓨터 시스템(1000)은 또한 통신 인터페이스(1024)를 포함할 수 있다. 통신 인터페이스(1024)는 소프트웨어 및 데이터가 컴퓨터 시스템(1000)과 외부 장치 사이에서 전송되도록 한다. 통신 인터페이스(1024)의 예는 모뎀, 네트워크 인터페이스 (이더넷 카드와 같은), 통신 포트, PCMCIA (Personal Computer Memory Card International Association) 슬롯 및 카드 등을 포함할 수 있다. 통신 인터페이스(1024)를 통해 전송 된 소프트웨어 및 데이터는 통신 인터페이스(1024)에 의해 수신 될 수 있는 전자, 전자기, 광학 또는 다른 신호 일 수 있는 신호(1028)의 형태이다. 이들 신호들(1028)은 통신 경로(1026)를 통해 통신 인터페이스(1024)에 제공된다. 통신 경로(1026)는 신호들(1028)을 전달하며, 유선 또는 케이블, 광섬유, 전화선, 무선 링크, 셀룰러 폰 링크, 무선 주파수 링크, 또는 임의의 다른 적절한 통신 채널을 사용하여 구현될 수 있다. 예를 들어, 통신 경로(1026)는 채널들의 조합을 사용하여 구현될 수 있다.
- [0081] 용어 "컴퓨터 프로그램 매체"및 "컴퓨터 사용 가능 매체"는 일반적으로 착탈식 저장 드라이브(1014), 하드 디스크 드라이브(1012)에 설치된 하드 디스크 및 신호(1028)와 같은 매체를 지칭하기 위해 사용된다. 이들 컴퓨터 프로그램 제품은 컴퓨터 시스템(1000)에 소프트웨어를 제공하기 위한 수단이다. 그러나 이 용어에는 여기에 설명 된 컴퓨터 프로그램을 구현하는 신호 (예: 전기, 광학 또는 전자기 신호)가 포함될 수도 있다.
- [0082] 컴퓨터 프로그램 (컴퓨터 제어 로직이라고도 함)은 주 메모리(1008) 및 / 또는 보조 메모리(1010)에 저장된다. 또한 컴퓨터 프로그램은 통신 인터페이스(1024)를 통해 수신될 수 있다. 이러한 컴퓨터 프로그램은 실행될 때 컴퓨터 시스템(1000)이 본 명세서에서 논의 된 바와 같이 본 발명의 실시예를 구현할 수 있게 한다. 따라서, 이러한 컴퓨터 프로그램은 컴퓨터 시스템(1000)의 제어기를 나타낸다. 실시예가 소프트웨어를 사용하여 구현되는 경우, 소프트웨어는 컴퓨터 프로그램 제품(1030)에 저장 될 수 있고 몇몇 예를 제공하기 위해 이동식 저장 드라이브(1014), 하드 디스크 드라이브(1012) 또는 통신 인터페이스(1024)를 사용하여 컴퓨터 시스템(1000)에 로드 될 수 있다.
- [0083] 다른 실시예들은 하드웨어, 펌웨어 또는 소프트웨어의 제어 로직 또는 이들의 임의의 조합으로서 구현될 수 있다. 예를 들어, 훈련된 CNN 모듈(11)은 이미지 프로세싱 시스템에서 컴포넌트로서 설치하기 위한 독립형 엔티티로서 하드웨어 및/또는 소프트웨어로 구현될 수 있으며, 트레이닝 모듈 기능 및/또는 깊이 데이터 생성기 기능을 더 포함할 수 있다.
- [0084] 본 발명의 실시예가 단지 예로서 설명되고 본 발명의 범위를 벗어나지 않고 다양한 변경 및 수정이 이루어질 수 있음을 이해할 것이다. 예를 들어, 상기 실시예들은 훈련된 통계 모델을 딥 컨볼루션 뉴럴 네트워크(deep convolutional neural network)로서 구현한다. 당업자라면 인식할 수 있는 바와 같이, 트레이닝 프로세스의 기본 양상은 랜덤 포레스트 및 과생 상품과 같은 예측 깊이 맵을 생성하기 위해 이미지 데이터를 처리하기에 적합한 다른 형태의 통계 모델에 적용 가능할 수 있다.
- [0085] 본 명세서에서 "일 실시예"는 반드시 동일한 실시예를 모두 지칭하는 것은 아니며, 다른 실시예와 상호 배타적인 별개의 또는 대안적인 실시예도 아니다. 특히, 전술한 실시예들의 양상들이 결합되어 추가적인 실시예들을

형성할 수 있음을 이해할 것이다. 유사하게, 일부 실시예에 의해 나타낼 수 있고 다른 것에는 나타내지 않을 수 없는 다양한 특징이 설명된다. 그럼에도 불구하고 다음의 청구 범위의 범주 내에 속하는 또 다른 대체 실시예가 구상될 수 있다.

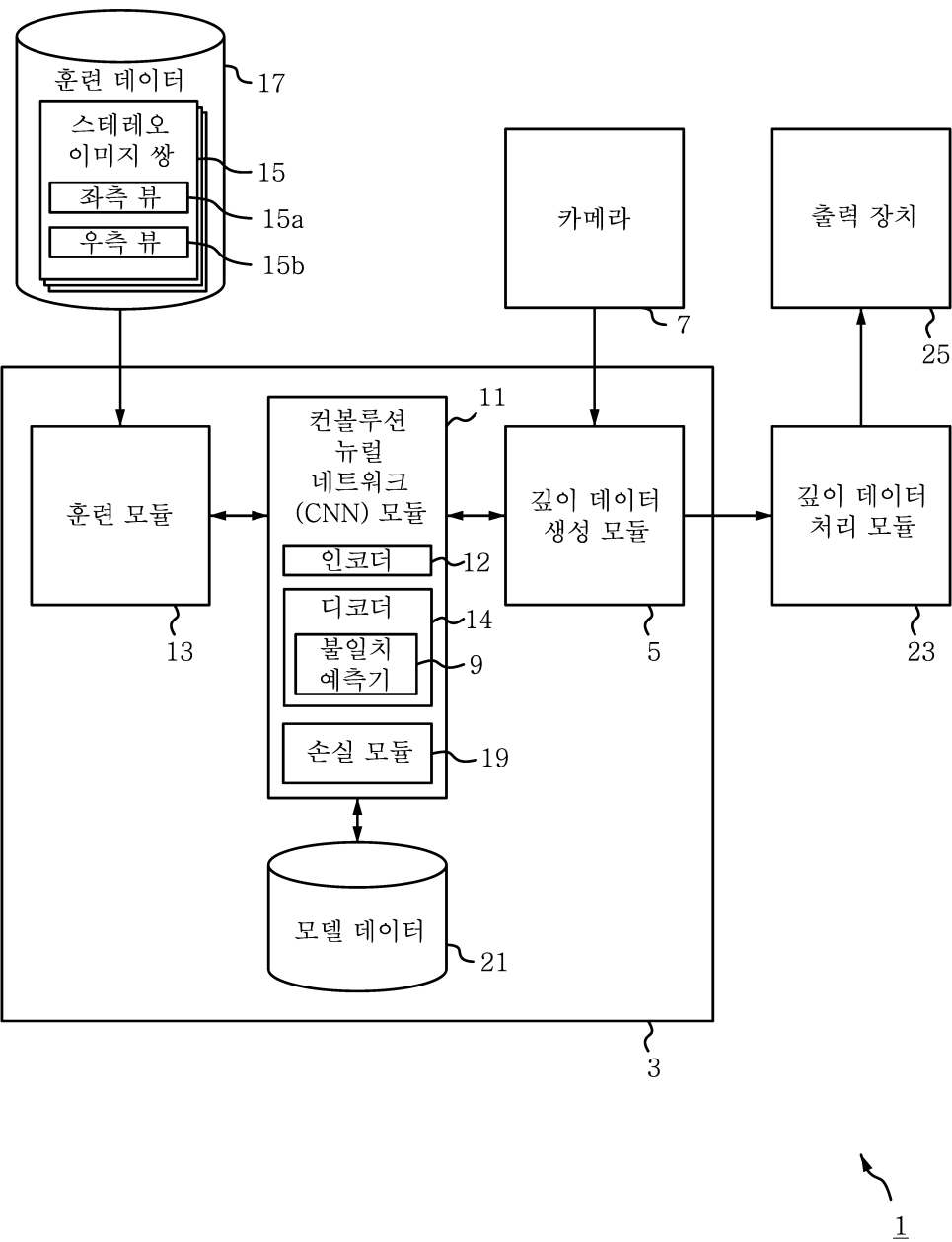
### 부호의 설명

[0086]

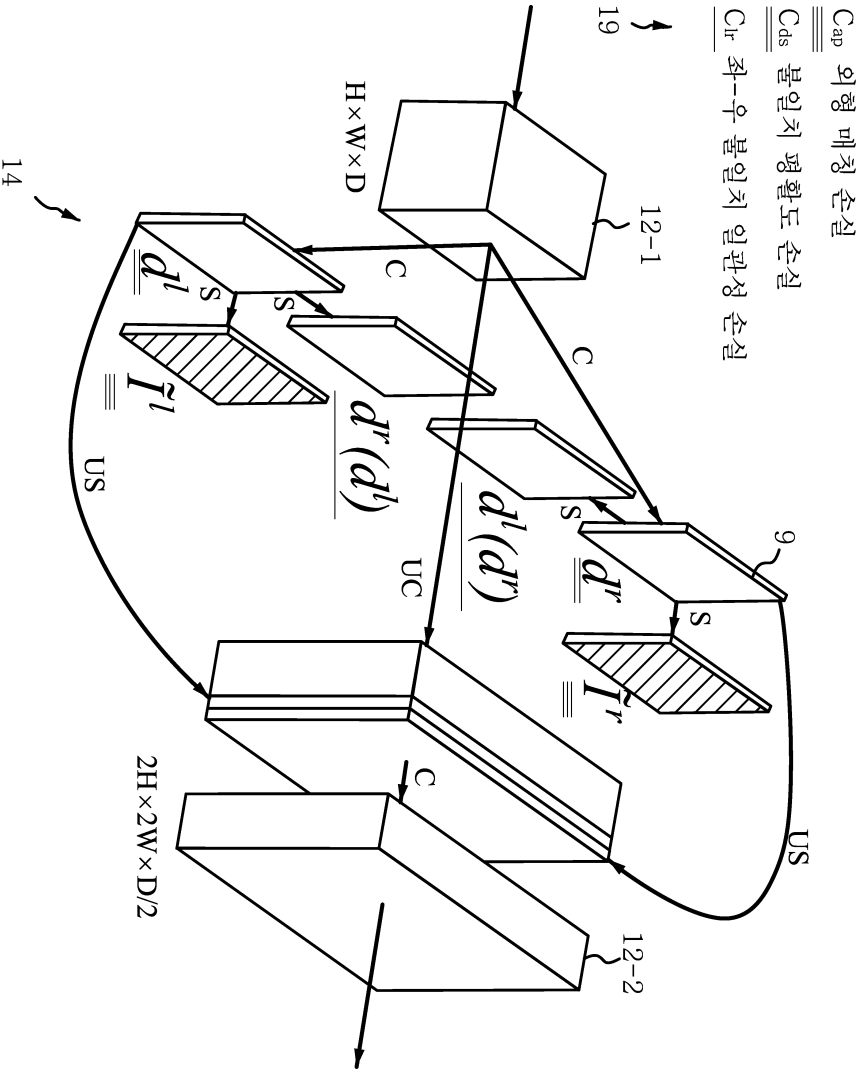
3 : 이미지 처리 시스템	5 : 깊이 데이터 생성 모듈
7 : 카메라	9 : 불일치 예측기
11: 컨볼루션 뉴럴 네트워크(CNN) 모듈	12 : 인코더
13: 훈련 모듈	14 : 디코더
15: 스테레오 이미지 쌍	15a : 좌측 뷰
15b : 우측 뷰	17: 훈련 데이터
19 : 손실 모듈	21 : 모델 데이터
23 : 깊이 데이터 처리 모듈	25 : 출력 장치

도면

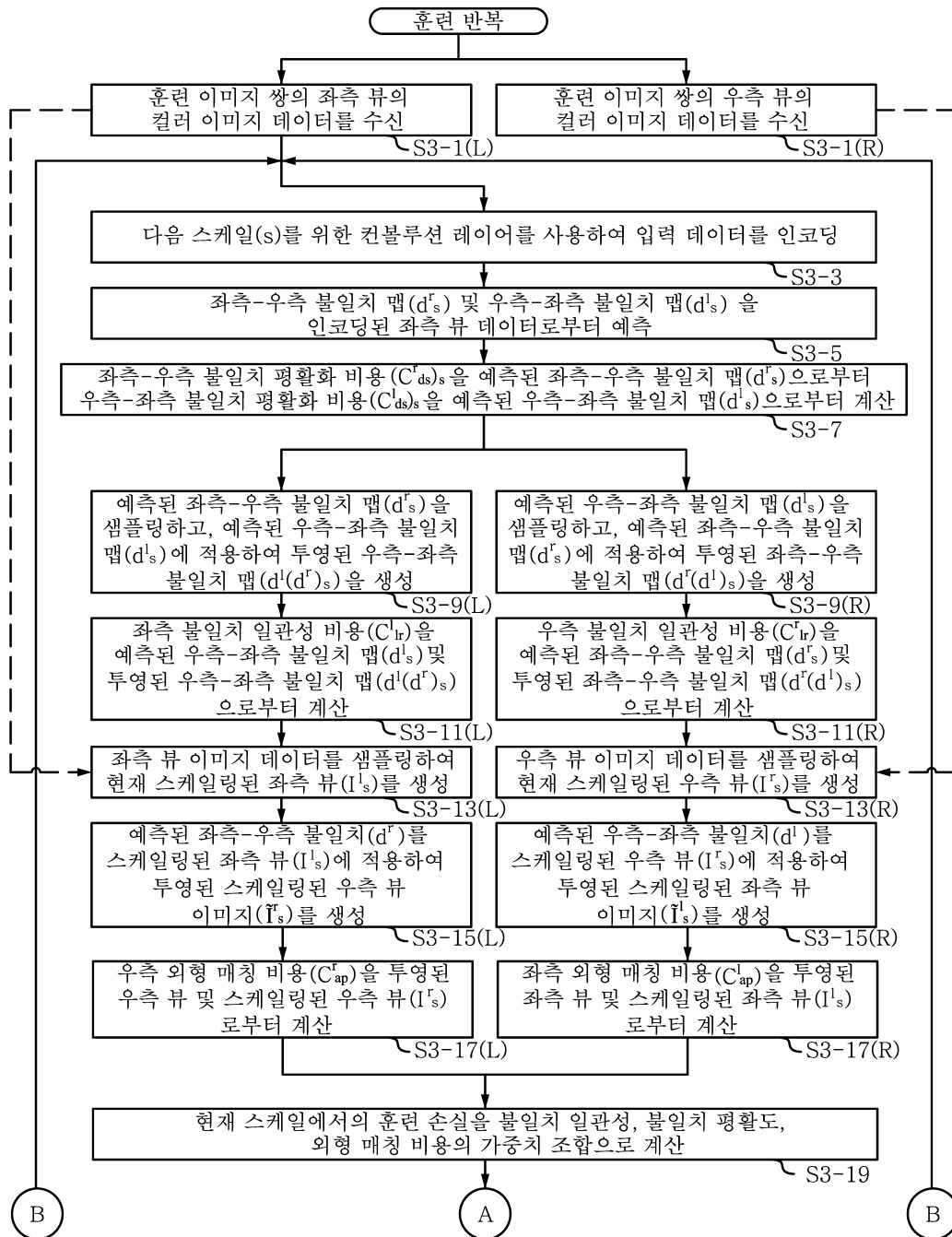
도면1



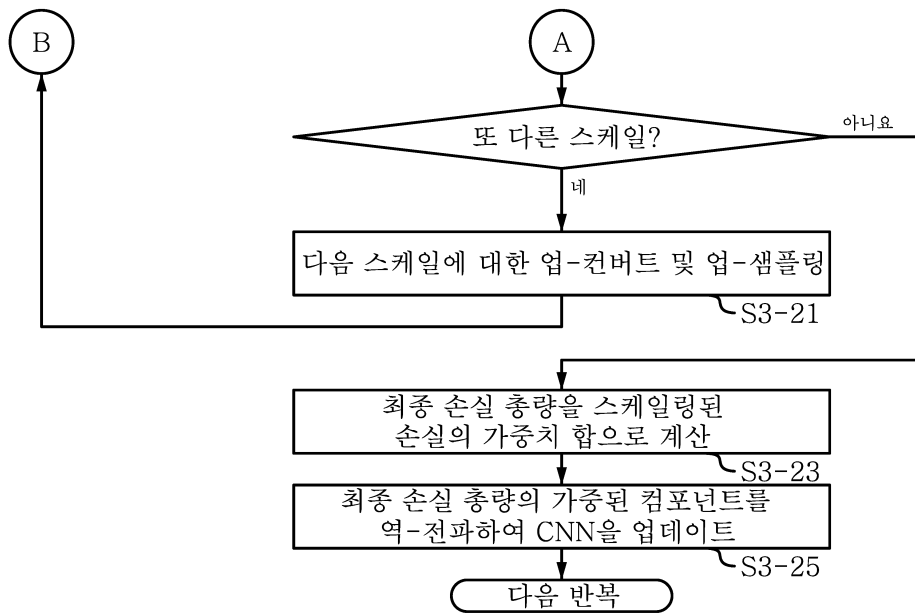
도면2



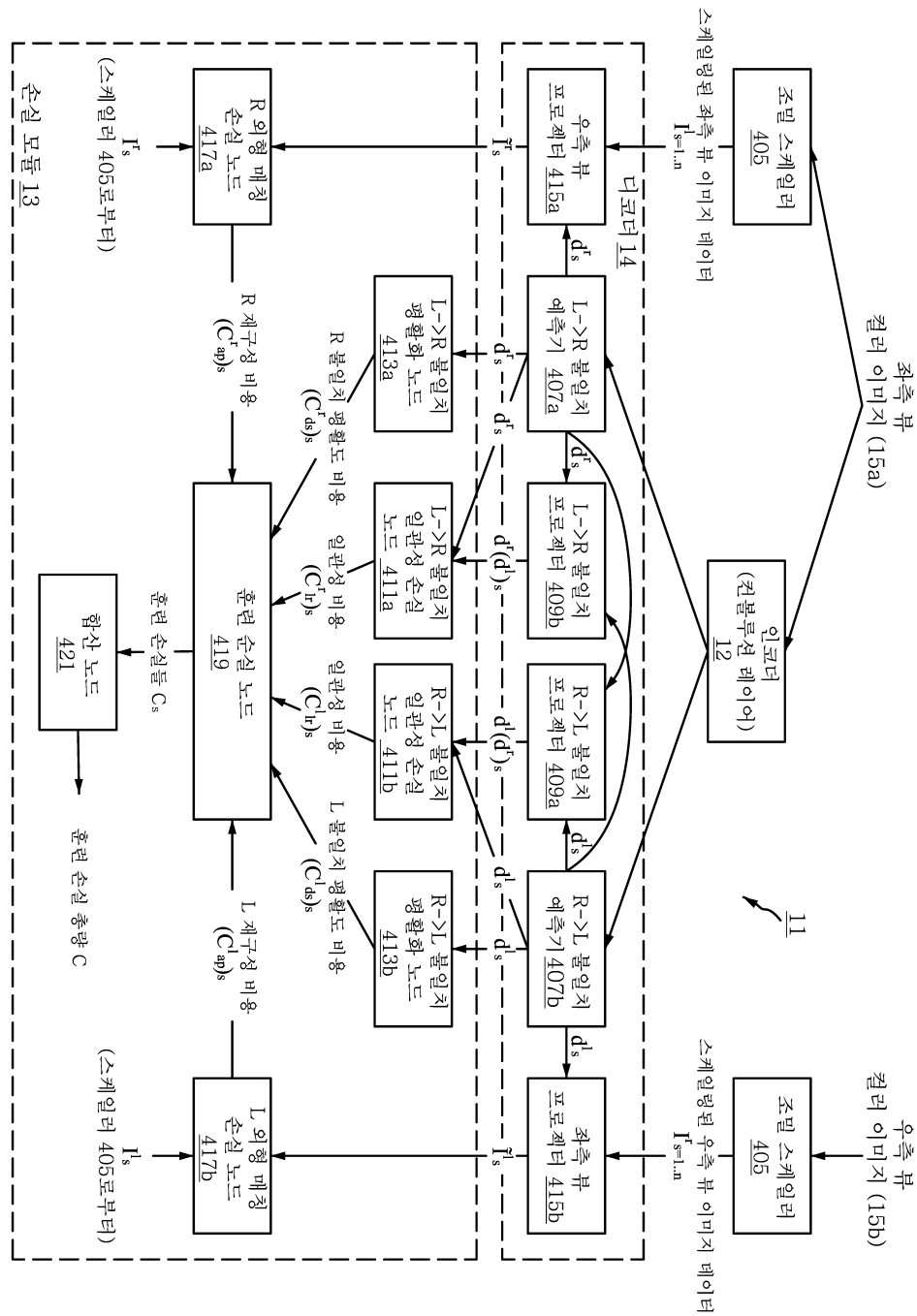
도면3a



도면3b

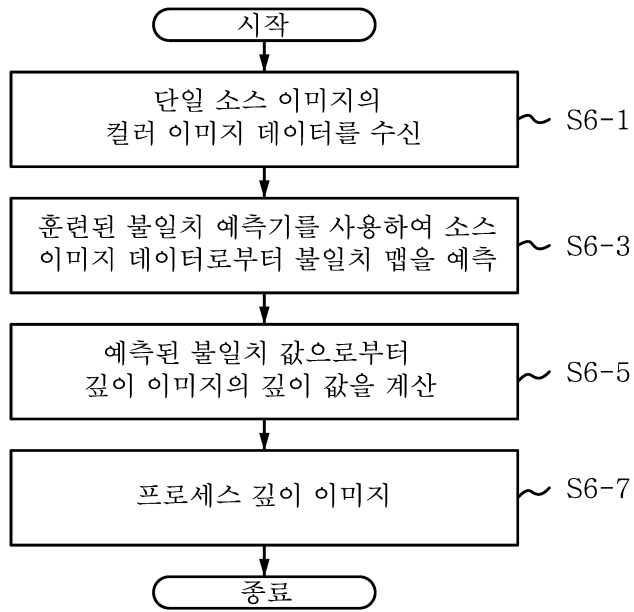


도면4





도면5



도면6

