



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0063368  
(43) 공개일자 2020년06월05일

(51) 국제특허분류(Int. Cl.)  
G06N 3/08 (2006.01) G06N 3/04 (2006.01)  
G06T 7/00 (2017.01) H04N 13/00 (2018.01)  
H04N 13/128 (2018.01)  
(52) CPC특허분류  
G06N 3/08 (2013.01)  
G06N 3/0454 (2013.01)  
(21) 출원번호 10-2018-0146709  
(22) 출원일자 2018년11월23일  
심사청구일자 2018년11월23일

(71) 출원인  
연세대학교 산학협력단  
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)  
(72) 발명자  
손광훈  
서울특별시 서대문구 연세로 50, 제3공학관 C129호(신촌동, 연세대학교)  
정성훈  
서울특별시 서대문구 연세로 50, 제3공학관 C129호(신촌동, 연세대학교)  
(74) 대리인  
민영준

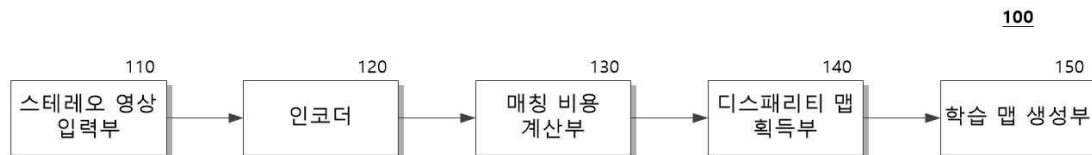
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 대응점 일관성에 기반한 비지도 학습 방식의 스테레오 매칭 장치 및 방법

(57) 요약

본 발명은 동일 구조와 동일한 가중치를 가지며 비지도 학습 방식으로 미리 학습된 2개의 컨볼루션 신경망(Convolutional Neural Networks: CNN)을 포함하여, 입력되는 스테레오 영상으로부터 특징 맵들을 추출하는 인코더, 특징 맵 사이의 매칭 비용 볼륨을 계산하는 매칭 비용 계산부 및 기지정된 최대 디스패리티 범위를 갖는 디스패리티 후보들 중 매칭 비용 볼륨을 최소화하는 디스패리티를 각 픽셀 별로 획득하고, 획득된 디스패리티로부터 디스패리티 맵을 생성하는 디스패리티 맵 획득부를 포함하고, 2개의 CNN은 학습 시에 입력된 스테레오 영상으로부터 획득된 디스패리티 맵에 대해 에피폴라 제약에 따른 대응점 일관성에 기반하여 양성 샘플을 추정하고, 추정된 양성 샘플을 인접 픽셀로 전파하여 생성되는 학습 맵들과 디스패리티 맵 사이의 오차를 역전파하여 학습되는 스테레오 매칭 장치 및 방법을 제공할 수 있다.

대표도



(52) CPC특허분류

**G06T 7/97** (2017.01)

**H04N 13/128** (2018.05)

**H04N 2013/0081** (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호 1711054882

부처명 과학기술정보통신부

연구관리전문기관 정보통신기술진흥센터(NIPA산하)

연구사업명 정보통신방송연구개발사업

연구과제명 [이지바로]스마트카 다중 센서와 딥러닝을 이용한 초정밀 내추럴 3D 뷰 생성 기술 개발(창조경제혁신센터 2단계)(2/2단계)(2/5연차)

기 여 율 1/1

주관기관 연세대학교 산학협력단

연구기간 2018.01.01 ~ 2018.12.31

---

## 명세서

### 청구범위

#### 청구항 1

동일 구조와 동일한 가중치를 가지며 비지도 학습 방식으로 미리 학습된 2개의 컨볼루션 신경망(Convolutional Neural Networks: CNN)을 포함하여, 입력되는 스테레오 영상으로부터 특징 맵들을 추출하는 인코더;

상기 특징 맵 사이의 매칭 비용 볼륨을 계산하는 매칭 비용 계산부; 및

기지정된 최대 디스패리티 범위를 갖는 디스패리티 후보들 중 매칭 비용 볼륨을 최소화하는 디스패리티를 각 픽셀 별로 획득하고, 획득된 디스패리티로부터 디스패리티 맵을 생성하는 디스패리티 맵 획득부; 를 포함하고,

상기 2개의 CNN은

학습 시에 입력된 스테레오 영상으로부터 획득된 디스패리티 맵에 대해 에피폴라 제약에 따른 대응점 일관성에 기반하여 양성 샘플을 추정하고, 추정된 양성 샘플을 인접 픽셀로 전파하여 생성되는 학습 맵들과 디스패리티 맵 사이의 오차를 역전파하여 학습되는 스테레오 매칭 장치.

#### 청구항 2

제1 항에 있어서, 상기 2개의 CNN은

디스패리티 맵에 대해 에피폴라 제약에 따른 대응점 일관성에 기반하여, 대응점 사이의 거리가 기지정된 임계값 미만인 픽셀을 회소 양성 샘플로 추정하여 회소 양성 샘플 맵을 획득하고,

상기 회소 양성 샘플 맵에서 상기 회소 양성 샘플을 보간 기법으로 인접 픽셀로 전파하여 보간 맵을 생성하며,

상기 보간 맵에 대해 다시 대응점 일관성따라 학습 샘플을 추정하며, 추정된 상기 학습 샘플로부터 상기 학습 맵을 생성하여 학습되는 스테레오 매칭 장치.

#### 청구항 3

제2 항에 있어서, 상기 보간 맵은

상기 회소 양성 샘플 맵에서 색상 유사성 제약 조건에 따라 입력 스테레오 영상의 컬러를 가이드로 이용하여, 상기 회소 양성 샘플을 인접 픽셀로 전파하여 생성되는 스테레오 매칭 장치.

#### 청구항 4

제3 항에 있어서, 상기 보간 맵은

회소 양성 샘플의 색상과 인접한 픽셀들 사이의 색상 평활도 제약에 따라 기지정된 에너지 함수가 최소화되는 보간 픽셀이 획득되고, 획득된 보간 픽셀로부터 생성되는 스테레오 매칭 장치.

#### 청구항 5

제1 항에 있어서, 상기 2개의 CNN은

학습 시에 입력된 디스패리티 맵과 상기 학습 맵 사이의 오차를 역전파하여 가중치를 업데이트 함으로써 학습되며, 상기 오차가 기지정된 기준 오차 이내가 되도록 반복적으로 학습 레이블을 생성하여 반복 학습되는 스테레오 매칭 장치.

#### 청구항 6

동일 구조와 동일한 가중치를 가지며 비지도 학습 방식으로 미리 학습된 2개의 컨볼루션 신경망(Convolutional Neural Networks: CNN)을 이용하여 입력되는 스테레오 영상으로부터 특징 맵들을 추출하는 단계;

상기 특징 맵 사이의 매칭 비용 볼륨을 계산하는 단계; 및

기지정된 최대 디스패리티 범위를 갖는 디스패리티 후보들 중 매칭 비용 볼륨을 최소화하는 디스패리티를 각 픽

셀 별로 획득하고, 획득된 디스패리티로부터 디스패리티 맵을 생성하는 단계; 를 포함하고,

상기 2개의 컨볼루션 신경망은

입력된 스테레오 영상으로부터 획득된 디스패리티 맵에 대해 에피폴라 제약에 따른 대응점 일관성에 기반하여 양성 샘플을 추정하고, 추정된 양성 샘플을 인접 픽셀로 전파하여 생성되는 학습 맵들과 디스패리티 맵 사이의 오차를 역전파하여 학습되는 스테레오 매칭 방법.

#### 청구항 7

제6 항에 있어서, 상기 특징 맵들을 추출하는 단계는

디스패리티 맵에 대해 에피폴라 제약에 따른 대응점 일관성에 기반하여, 대응점 사이의 거리가 기지정된 임계값 미만인 픽셀을 최소 양성 샘플로 추정하여 최소 양성 샘플 맵을 획득하는 단계;

상기 최소 양성 샘플 맵에서 상기 최소 양성 샘플을 보간 기법으로 인접 픽셀로 전파하여 보간 맵을 생성하는 단계; 및

상기 보간 맵에 대해 다시 대응점 일관성 따라 학습 샘플을 추정하며, 추정된 상기 학습 샘플로부터 상기 학습 맵을 생성하는 단계; 를 포함하는 학습 방법에 따라 미리 학습되는 스테레오 매칭 방법.

#### 청구항 8

제7 항에 있어서, 상기 보간 맵을 생성하는 단계는

상기 최소 양성 샘플 맵에서 색상 유사성 제약 조건에 따라 입력 스테레오 영상의 컬러를 가이드로 이용하여, 상기 최소 양성 샘플을 인접 픽셀로 전파하여 생성되는 스테레오 매칭 방법.

#### 청구항 9

제8 항에 있어서, 상기 보간 맵을 생성하는 단계는

최소 양성 샘플의 색상과 인접한 픽셀들 사이의 색상 평활도 제약에 따라 기지정된 에너지 함수가 최소화되는 보간 픽셀이 획득하는 단계; 및

획득된 보간 픽셀을 이용하여 상기 보간 맵을 생성하는 단계; 를 포함하는 스테레오 매칭 방법.

#### 청구항 10

제6 항에 있어서, 상기 특징 맵들을 추출하는 단계는

학습 시에 입력된 디스패리티 맵과 상기 학습 맵 사이의 오차를 역전파하여 가중치를 업데이트 함으로써 학습되며, 상기 오차가 기지정된 기준 오차 이내가 되도록 반복적으로 학습 레이블을 생성하여 반복 학습되는 스테레오 매칭 방법.

### 발명의 설명

### 기술 분야

[0001] 본 발명은 스테레오 매칭 장치 및 방법에 관한 것으로, 대응점 일관성에 기반한 비지도 학습 방식의 스테레오 매칭 장치 및 방법에 관한 것이다.

### 배경 기술

[0002] 스테레오 매칭은 영상으로부터 3차원 기하학적 구성을 인식하기 위한 방법으로, 컴퓨터 비전 시스템의 스테레오 영상 재구성, 자율 주행, 운전자 보조 시스템(Advanced Driver Assistance System: 이하 ADAS), 로봇 공학 등을 포함하는 다양한 분야에 이용되고 있다.

[0003] 스테레오 매칭은 서로 다른 2개의 시점 영상을 갖는 스테레오 영상에서 3차원 위치 정보(깊이 정보)를 추정하는 기법으로, 스테레오 영상에서 대응점(correspondence) 사이의 비유사성을 측정하는 매칭 비용(matching cost) 계산은 스테레오 매칭 기법의 핵심 과정이다.

- [0004] 그러나 매칭 비용 계산은 스테레오 영상의 폐색(occlusion) 영역, 질감없는(textureless) 영역 또는 조명의 변화 등과 같은 영상 자체의 모호성(inherent matching ambiguity)으로 인해 작업에 어려움이 있다.
- [0005] 이러한 매칭 비용 계산의 어려움을 극복하기 위하여 다양한 기법이 제안되었으며, 최근에는 딥 러닝(Deep learning) 기법으로 학습된 인공 신경망(Artificial Neural Network)을 이용하여 영상으로부터 매칭 비용을 계산하는 방법이 제안되었다. 특히 인공 신경망 중 영상 처리 분야에서 탁월한 성능을 나타내는 컨볼루션 신경망(Convolution Neural Network)이 매칭 비용 계산에 주로 이용되고 있다.
- [0006] 매칭 비용 계산에 인공 신경망을 이용하는 경우, 인공 신경망은 스테레오 영상에서 대응점 간의 매칭 비용을 최소화할 수 있도록 미리 학습되어야 한다. 그러나 인공 신경망이 신뢰성 있게 매칭 비용을 계산할 수 있도록 학습되기 위해서는 학습 레이블이 포함된 대량의 학습 영상을 요구하는 지도 학습(supervised learning)이 수행되어야 한다.
- [0007] 학습 레이블은 구조형 조명(structured light)이나 라이다(Light Detection And Ranging) 및 레이저 스캐너 등과 같은 3D 센서를 이용하여 직접 검증(ground truth)된 깊이 값을 측정함으로써 획득될 수 있다. 그러나 3D 센서를 이용한 측정은 고비용을 요구될 뿐만 아니라, 사용 조건에 따라 다양한 오류가 존재하여 수작업에 의한 오류 보정이 추가로 요구되는 문제가 있다.
- [0008] 이에 학습을 위해 학습 레이블이 포함되어 제공되는 학습 영상의 개수가 부족하여, 인공 신경망이 정상적으로 학습되지 않아 매칭 비용을 정상적으로 추정할 수 없는 문제가 있다.

## 선행기술문헌

### 특허문헌

- [0009] (특허문헌 0001) 한국 등록 특허 제10-1354387호 (2014.01.15 등록)

## 발명의 내용

### 해결하려는 과제

- [0010] 본 발명의 목적은 학습 레이블을 필요로 하지 않는 비지도 학습 방식으로 학습된 인공 신경망을 이용하는 스테레오 매칭 장치 및 방법을 제공하는데 있다.
- [0011] 본 발명의 다른 목적은 입력된 스테레오 영상 사이의 에피폴라 제약과 대응 일관성을 이용하여 매칭 비용을 최소화하도록 학습된 인공 신경망을 이용하는 스테레오 매칭 장치 및 방법을 제공하는데 있다.
- [0012] 본 발명의 또 다른 목적은 조명 변화 및 야외 환경과 같은 다양한 환경에서 신뢰성 있는 스테레오 매칭을 수행할 수 있는 스테레오 매칭 장치 및 방법을 제공하는데 있다.

### 과제의 해결 수단

- [0013] 상기 목적을 달성하기 위한 본 발명의 일 실시예에 따른 스테레오 매칭 장치는 동일 구조와 동일한 가중치를 가지며 비지도 학습 방식으로 미리 학습된 2개의 컨볼루션 신경망(Convolutional Neural Networks: CNN)을 포함하여, 입력되는 스테레오 영상으로부터 특징 맵들을 추출하는 인코더; 상기 특징 맵 사이의 매칭 비용 볼륨을 계산하는 매칭 비용 계산부; 및 기지정된 최대 디스패리티 범위를 갖는 디스패리티 후보들 중 매칭 비용 볼륨을 최소화하는 디스패리티를 각 픽셀 별로 획득하고, 획득된 디스패리티로부터 디스패리티 맵을 생성하는 상기 디스패리티 맵 획득부;를 포함하고, 상기 2개의 CNN은 학습 시에 입력된 스테레오 영상으로부터 획득된 디스패리티 맵에 대해 에피폴라 제약에 따른 대응점 일관성에 기반하여 양성 샘플을 추정하고, 추정된 양성 샘플을 인접 픽셀로 전파하여 생성되는 학습 맵들과 디스패리티 맵 사이의 오차를 역전파하여 학습된다.
- [0014] 상기 2개의 CNN은 디스패리티 맵에 대해 에피폴라 제약에 따른 대응점 일관성에 기반하여, 대응점 사이의 거리가 기지정된 임계값 미만인 픽셀을 최소 양성 샘플로 추정하여 최소 양성 샘플 맵을 획득하고, 상기 최소 양성 샘플 맵에서 상기 최소 양성 샘플을 보간 기법으로 인접 픽셀로 전파하여 보간 맵을 생성하며, 상기 보간 맵에 대해 다시 대응점 일관성따라 학습 샘플을 추정하며, 추정된 상기 학습 샘플로부터 상기 학습 맵을 생성하여 학습될 수 있다.

[0015] 상기 보간 맵은 상기 희소 양성 샘플 맵에서 색상 유사성 제약 조건에 따라 입력 스테레오 영상의 컬러를 가이드로 이용하여, 상기 희소 양성 샘플을 인접 픽셀로 전파하여 생성될 수 있다.

[0016] 상기 2개의 CNN은 학습 시에 입력된 디스패리티 맵과 상기 학습 맵 사이의 오차를 역전파하여 가중치를 업데이트 함으로써 학습되며, 상기 오차가 기지정된 기준 오차 이내가 되도록 반복적으로 학습 레이블을 생성하여 반복 학습될 수 있다.

[0017] 상기 목적을 달성하기 위한 본 발명의 다른 실시예에 따른 스테레오 매칭 방법은 동일 구조와 동일한 가중치를 가지며 비지도 학습 방식으로 미리 학습된 2개의 컨볼루션 신경망(Convolutional Neural Networks: CNN)을 이용하여 입력되는 스테레오 영상으로부터 특징 맵들을 추출하는 단계; 상기 특징 맵 사이의 매칭 비용 볼륨을 계산하는 단계; 및 기지정된 최대 디스패리티 범위를 갖는 디스패리티 후보들 중 매칭 비용 볼륨을 최소화하는 디스패리티를 각 픽셀 별로 획득하고, 획득된 디스패리티로부터 디스패리티 맵을 생성하는 단계; 를 포함하고, 상기 2개의 컨볼루션 신경망은 입력된 스테레오 영상으로부터 획득된 디스패리티 맵에 대해 에피폴라 제약에 따른 대응점 일관성에 기반하여 양성 샘플을 추정하고, 추정된 양성 샘플을 인접 픽셀로 전파하여 생성되는 학습 맵들과 디스패리티 맵 사이의 오차를 역전파하여 학습된다.

### 발명의 효과

[0018] 따라서, 본 발명의 실시예에 따른 스테레오 매칭 장치 및 방법은 입력된 스테레오 영상 사이의 에피폴라 제약과 대응점 일관성을 이용하여 양성 샘플을 추정하고, 추정된 양성 샘플을 전파하여 학습 샘플 맵을 획득함으로써, 비지도 학습 방법으로 학습될 수 있다. 또한 조명 변화나 야외 환경과 같은 다양한 환경에서도 신뢰성 있는 스테레오 매칭을 수행할 수 있다.

### 도면의 간단한 설명

[0019] 도1 은 본 발명의 일 실시예에 따른 스테레오 매칭 장치의 개략적 구성을 나타낸다.

도2 는 도1 의 학습 맵 생성부의 상세 구성을 나타낸다.

도3 은 도1 의 스테레오 매칭 장치의 각 구성별 동작을 설명하기 위한 도면이다.

도4 는 본 발명의 실시예에 따라 생성된 학습 맵의 예를 나타낸다.

도5 는 본 발명의 일 실시예에 따른 스테레오 매칭 방법 및 이의 학습 방법에 대한 개략적 구성을 나타낸다.

도6 는 도5 의 학습 맵 생성 및 학습 단계를 상세하게 나타낸다.

도7 은 도6 의 학습 맵 생성 및 학습 단계의 알고리즘을 나타낸다.

도8 은 본 발명의 실시예에 따른 학습 맵의 생성 과정에서 생성되는 맵의 예를 나타낸다.

도9 는 본 발명의 실시예에 따른 학습 샘플과 다른 비지도 학습 방식으로 획득되는 샘플을 비교한 도면이다.

도10 내지 도12 는 본 발명의 실시예에 따른 CNN의 구조에 따른 학습 샘플을 비교한 도면이다.

도13 및 도14 는 본 발명의 실시예에 따른 CNN의 구조에 따른 스테레오 매칭 결과를 비교한 도면이다.

도15 및 도16 은 본 실시예에 따른 비지도 학습 방법의 조명에 대한 강건성을 시뮬레이션한 결과를 나타낸다.

도17 및 도18 은 야외 운전 환경에서의 스테레오 매칭 성능을 시뮬레이션한 결과를 나타낸다.

### 발명을 실시하기 위한 구체적인 내용

[0020] 본 발명과 본 발명의 동작상의 이점 및 본 발명의 실시예에 의하여 달성되는 목적을 충분히 이해하기 위해서는 본 발명의 바람직한 실시예를 예시하는 첨부 도면 및 첨부 도면에 기재된 내용을 참조하여야만 한다.

[0021] 이하, 첨부한 도면을 참조하여 본 발명의 바람직한 실시예를 설명함으로써, 본 발명을 상세히 설명한다. 그러나, 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며, 설명하는 실시예에 한정되는 것이 아니다. 그리고, 본 발명을 명확하게 설명하기 위하여 설명과 관계없는 부분은 생략되며, 도면의 동일한 참조부호는 동일한 부재임을 나타낸다.

[0022] 명세서 전체에서, 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다

른 구성요소를 제외하는 것이 아니라, 다른 구성요소를 더 포함할 수 있는 것을 의미한다. 또한, 명세서에 기재된 "...부", "...기", "모듈", "블록" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.

[0023] 도1 은 본 발명의 일 실시예에 따른 스테레오 매칭 장치의 개략적 구성을 나타내고, 도2 는 도1 의 학습 맵 생성부의 상세 구성을 나타내며, 도3 은 도1 의 스테레오 매칭 장치의 각 구성별 동작을 설명하기 위한 도면이다.

[0024] 도1 을 참조하면, 본 실시예에 따른 스테레오 매칭 장치(100)는 스테레오 영상 입력부(110), 인코더(120), 매칭 비용 계산부(130), 디스패리티 맵 획득부(140) 및 학습 맵 생성부(150)를 포함한다.

[0025] 스테레오 영상 입력부(110)는 스테레오 매칭을 수행해야할 스테레오 영상을 획득한다. 여기서 스테레오 영상은 스테레오 카메라에서 획득될 수 있는 영상으로 서로 다른 시점을 갖는 2개의 영상으로 구성될 수 있다. 스테레오 영상은 스테레오 카메라의 구조에 따라 상하 영상 좌우 영상 등으로 획득될 수 있으나, 여기서는 일예로 도3 에 도시된 바와 같이 좌영상(Left image)( $I^l$ )과 우영상(Right image)( $I^r$ )을 획득하는 것으로 가정한다.

[0026] 이에 인코더(120)는 스테레오 영상 입력부(110)에서 획득된 좌영상( $I^l$ )과 우영상( $I^r$ )을 인코딩하여 매칭 비용을 최소화하도록 하는 2개의 특징 맵( $A^l, A^r$ )을 획득한다.

[0027] 인코더(120)는 도3 에 도시된 바와 같이 미리 학습된 2개의 컨볼루션 신경망(Convolution neural network: 이하 CNN)을 포함하며, 인코더(120)에 포함된 2개의 CNN은 동일한 구조를 갖고, 동일한 가중치(w)가 적용되어 피드 포워드 프로세스를 수행하는 삼(siamese) CNN으로 구현된다. 인코더(120)의 2개의 CNN은 지정된 패턴 인식 기법에 따라 미리 학습됨으로써, 입력 영상인 좌영상( $I^l$ )과 우영상( $I^r$ ) 각각으로부터 좌 특징 맵( $A^l$ )과 우 특징 맵( $A^r$ )을 획득한다.

[0028] 인코더(120)에 포함된 2개의 CNN이 동일한 가중치(w)가 적용되어 피드 포워드 프로세스(Fw)를 수행하므로, 2개의 특징 맵( $A^l, A^r$ )의 각 픽셀(i)은 수학식 1과 같이 획득된다.

### 수학식 1

$$A^l(i) = \mathcal{F}_W(I^l(i))$$

$$A^r(i) = \mathcal{F}_W(I^r(i))$$

[0029]

[0030] 매칭 비용 계산부(130)는 인코더(120)에서 출력된 특징 맵( $A^l, A^r$ )을 인가받아 특징 맵( $A^l, A^r$ )들 사이의 매칭 비용 볼륨(matching cost volume)( $C^l, C^r$ )을 수학식 2와 같이 계산한다.

### 수학식 2

$$C^l(i, d) = \|A^l(i_x, i_y) - A^r(i_x - d, i_y)\|_1$$

[0031]

[0032] (여기서,  $\|\cdot\|_1$ 은  $l_1$ -norm 함수를 나타낸다.)

[0033] 수학식 2는 좌 특징 맵( $A^l$ )에 대한 좌 매칭 비용 볼륨( $C^l$ )을 계산하는 수학식을 나타내었으나, 좌 특징 맵( $A^r$ )에 대한 좌 매칭 비용 볼륨( $C^r$ )도 유사하게 계산될 수 있다.

[0034] 디스패리티 맵 획득부(140)는 매칭 비용 계산부(130)에서 매칭 비용 볼륨( $C^l, C^r$ )이 계산되면, 모든 픽셀(i)에 대해 최대 디스패리티 범위(maximum disparity range)( $d_{\max}$ )를 갖는 디스패리티 후보들( $d=\{1, \dots, d_{\max}\}$ ) 중 매칭 비용 볼륨( $C^l, C^r$ )이 최소화되는 디스패리티(d)를 수학식 3에 따라 탐색하여 획득함으로써, 디스패리티 맵을



생성한다.

### 수학식 3

$$d^l(i) = \operatorname{argmin}_k C^l(i, k)$$

[0035]

[0036]

수학식 3에서도 예시로서 좌 디스패리티( $d^l$ )를 계산하는 수학식만을 개시하였으나, 우 디스패리티( $d^r$ )에 대해서도 유사하게 계산할 수 있다.

[0037]

수학식 3에 따르면, 디스패리티 맵 획득부(140)는 WTA(winner-takes-all) 방식으로 디스패리티 검색 범위에서 최소 매칭 비용을 갖는 디스패리티( $d$ )를 획득한다.

[0038]

즉 본 실시예에 따른 스테레오 매칭 장치(100)는 동일한 구조와 동일한 가중치( $w$ )를 갖는 삼 CNN을 이용함으로써, 입력된 스테레오 영상( $I^l, I^r$ )으로부터  $I^l(i_x, i_y) = I^r(i_x - d^l(i), i_y)$ 를 만족하는 밀집 대응점(dense correspondence)을 확인하여, 각 픽셀( $i$ )에 대한 좌 디스패리티( $d^l(i)$ )를 추정한다.

[0039]

이때 스테레오 매칭 장치(100)는 최대 디스패리티 범위( $d_{\max}$ )를 갖는 디스패리티 후보( $d = \{1, \dots, d_{\max}\}$ ) 중에서 적합한 좌 디스패리티( $d^l(i)$ )를 추정하기 위해,  $I^l(i_x, i_y)$ 과  $I^r(i_x - d, i_y)$  사이의 매칭 비용( $C^l(i)$ )을 계산한다. 여기서 좌 디스패리티( $d^l(i)$ )는 WTA 기법에 따라 매칭 비용( $C^l(i)$ )을 최소화하도록 결정될 수 있다. 그리고 획득된 좌 디스패리티( $d^l(i)$ )로부터 좌 디스패리티 맵( $d^l$ )을 획득할 수 있다.

[0040]

우 디스패리티 맵( $d^r$ ) 또한 유사한 방식으로 획득될 수 있다.

[0041]

상기에서는 스테레오 매칭 장치(100)가 좌 디스패리티 맵( $d^l$ )과 우 디스패리티 맵( $d^r$ )을 모두 획득하는 것으로 설명하였으나, 노이즈가 포함되지 않은 스테레오 영상에 대해 정확하게 좌 디스패리티 맵( $d^l$ )과 우 디스패리티 맵( $d^r$ )이 획득된다면, 좌 디스패리티 맵( $d^l$ )과 우 디스패리티 맵( $d^r$ )은 상호 대칭 구조를 갖게 된다. 따라서 스테레오 매칭 장치(100)는 좌 디스패리티 맵( $d^l$ )과 우 디스패리티 맵( $d^r$ ) 중 적어도 하나를 스테레오 매칭 결과인 디스패리티 맵( $d$ )으로서 출력할 수도 있다.

[0042]

한편, CNN을 이용하여 매칭 비용을 계산하기 위해서는 상기한 바와 같이, CNN이 미리 학습되어야 하며, 이에 학습 레이블이 포함된 대량의 학습 영상이 요구된다. 그러나 학습 레이블이 포함된 대량의 학습 영상은 획득하기 어려우며, 따라서 CNN을 학습시키는 것이 용이하지 않다.

[0043]

이에 본 실시예에서는 인코더(120)의 CNN 학습 시에 학습 레이블이 포함된 학습 영상을 요구하지 않고, 학습 레이블이 포함되지 않은 스테레오 영상을 이용하여 학습 레이블을 생성할 수 있도록 학습 맵 생성부(150)를 추가로 제안한다.

[0044]

학습 맵 생성부(150)는 인코더(120)의 삼 CNN을 학습시키기 위해 추가되는 학습 레이블(또는 학습 레이블이 포함된 학습 영상)을 생성하는 구성이다. 즉 본 실시예에 따른 스테레오 매칭 장치(100)를 학습 시키기 위한 학습 장치에 포함되는 구성으로, 인코더(120)의 삼 CNN이 이미 학습된 경우, 제외될 수 있다.

[0045]

학습 시에 스테레오 영상 입력부(110)는 인코더(120)의 삼 CNN을 학습시키기 위해 다수개의 스테레오 영상을 획득할 수 있다. 여기서는 일예로 스테레오 영상 입력부(110)가 T(여기서 T는 자연수)개 스테레오 영상을 획득하는 것으로 가정한다.

[0046]

그리고 학습시에 인코더(120)의 삼 CNN에 대한 가중치( $w$ )의 초기값은 기지정된 범위(예를 들면,  $0 < w \leq 1$ )에서 랜덤하게 선택된다.

[0047]

학습 시에는 인코더(120)의 삼 CNN이 정상적으로 기능하지 않으므로, 인코더(120)에서 출력되는 특징 맵( $A^l, A^r$ )으로부터 획득되는 매칭 비용 볼륨( $C$ )과 디스패리티 맵( $d^l, d^r$ ) 또한 신뢰할 수 없다.



- [0048] 그러나 본 발명의 학습 맵 생성부(150)는 학습되지 않은 인코더(120)에서 출력된 특징 맵( $A^l$ ,  $A^r$ )으로부터 획득된 디스패리티 맵( $d^l$ ,  $d^r$ )에 대해 에피폴라 제약(epipolar constraint)에 따른 대응점 일관성(Correspondence Consistency)에 기반하여 신뢰할 수 있는 양성 샘플을 추정하고, 추정된 양성 샘플을 전파함으로써, 인코더(120)의 삼 CNN을 학습 시킬 수 있는 학습 맵을 생성한다.
- [0049] 학습 맵 생성부(150)는 동일한 입력 스테레오 영상에 대해 반복적으로 학습 맵을 생성하여, 학습 맵의 신뢰도를 점차로 증가시킨다. 그리고 다수개 스테레오 영상 각각에 대한 학습 맵을 생성함으로써, 삼 CNN이 다양한 스테레오 영상에 대해 매칭 비용을 최소화하기 위한 특징 맵( $A^l$ ,  $A^r$ )을 신뢰성 있게 추출할 수 있도록 한다.
- [0050] 도3 을 참조하면, 학습 맵 생성부(150)는 양성 샘플 추출부(151), 보간 맵 생성부(153) 및 학습 맵 획득부(155)를 포함할 수 있다.
- [0051] 양성 샘플 추출부(151)는 디스패리티 맵 획득부(140)에서 획득된 좌 디스패리티 맵( $d^l$ ) 및 우 디스패리티 맵( $d^r$ )을 인가받고, 좌 디스패리티 맵( $d^l$ ) 및 우 디스패리티 맵( $d^r$ )의 각 디스패리티( $d^l(i)$ ,  $d^r(i)$ )에 대해 에피폴라 제약에 따른 대응점 일관성에 기반하여 양성 샘플을 탐색하여 추출한다.
- [0052] 스테레오 매칭에서 좌 영상의 픽셀들은 에피폴라 라인(epipolar line)을 가로질러 우 영상에 최대 하나의 매칭 픽셀을 가지며, 반대의 경우도 동일하다. 즉 좌 영상에서 우 영상으로의 픽셀 일치는 우 영상의 대응점(픽셀)이 좌영상에도 일치해야 한다.
- [0053] 이러한 에피폴라 제약에 따른 대응점 일관성을 이용하면, 수학적 4와 같이 좌 디스패리티 맵( $d^l$ ) 및 우 디스패리티 맵( $d^r$ )에서 신뢰할 수 있는 양성 샘플을 추출할 수 있다.

#### 수학적 4

[0054] 
$$|d^l(i_x, i_y) - d^r(i_x - d^l(i_x, i_y), i_y)| < t$$

- [0055] 여기서 t는 기지정된 임계값이다.
- [0056] 즉 수학적 4 에 따르면, 좌 디스패리티 맵( $d^l$ )과 우 디스패리티 맵( $d^r$ )에서 대응점 사이의 거리가 임계값(t) 미만인 픽셀이 희소 양성 샘플(sparse positive sample)( $\hat{d}_i$ )로 추정된다.
- [0057] 그리고 양성 샘플 추출부(151)는 추정된 희소 양성 샘플( $\hat{d}_i$ ) 이외의 나머지 픽셀은 일관성이 없는 불량 픽셀로 판단하여 제거한다.
- [0058] 이때 추정된 희소 양성 샘플( $\hat{d}_i$ )은 모든 영상 도메인에 분포될 수 있으나, 객체의 경계와 질감없는(textureless) 영역의 모호함으로 인해 오류 샘플이 포함될 수 있으며, 이로 인해 학습 성능 저하를 유발할 수 있다.
- [0059] 이에 양성 샘플 추출부(151)는 대응점 일관성에 의해 좌 디스패리티 맵( $d^l$ )과 우 디스패리티 맵( $d^r$ ) 각각 에서 추정된 양성 샘플은 유사한 색상값을 갖는다는 색상 유사성 제약 조건을 추가로 적용할 수 있다. 이 경우, 추정된 양성 샘플에서 색상 차가 기지정된 기준값 미만인 양성 샘플 만을 희소 양성 샘플( $\hat{d}_i$ )로 추정할 수 있다.
- [0060] 그리고 보간 맵 생성부(153)는 희소 양성 샘플( $\hat{d}_i$ )을 주변 픽셀로 전파하여 보간 맵(p)을 생성한다. 보간 맵 생성부(153)는 입력 컬러 영상을 가이드로 이용하여 희소 양성 샘플( $\hat{d}_i$ )을 인접 픽셀로 전파하는 보간 기법을 적용한다.
- [0061] 보간 맵 생성부(153)는 반복되는 학습 과정에서 수학적 5에 따른 전체 에너지 함수(J(p))가 최소화되도록 보간

을 수행하여 보간 픽셀( $p_i$ )을 획득한다.

### 수학식 5

$$J(p) = \sum_i \left( h_i(p_i - \hat{d}_i)^2 + \lambda \sum_{j \in \mathcal{N}_4(i)} w_{i,j}(I) (p_i - p_j)^2 \right)$$

여기서  $\lambda$ 는 두 디스패리티 맵( $d^l, d^r$ )에서 추정된 최소 양성 샘플( $\hat{d}_i$ ) 사이의 밸런스를 제어하기 위한 상수이며,  $h_i$ 는 유효 픽셀인 경우 1이고 아니면 0을 나타내는 색인 함수이며,  $\mathcal{N}_4(i)$ 는 보간 픽셀( $p_i$ )에 대한 인접 픽셀의 집합을 나타낸다. 그리고 범위 파라미터( $\sigma_c$ )와 함께  $w_{i,j}(I) = \exp(-\|I_i - I_j\|/\sigma_c)$ 로 정의되는 공간 변이 가중치 함수( $w_{i,j}(I)$ )를 이용하여 평활도 제약이 적응적으로 강제된다.

그리고 보간 픽셀( $p_i$ )로 획득된 보간된 디스패리티 맵인 보간 맵( $p$ )은 수학식 6에 의해 계산될 수 있다.

### 수학식 6

$$p(m) = \frac{((I + \lambda L)^{-1} \hat{d})(m)}{((I + \lambda L)^{-1} h)(m)}$$

여기서  $p$ 와  $\hat{d}$ 은 픽셀 수  $S$ 를 갖는 보간 맵( $p$ )과 최소 양성 샘플 맵( $\hat{d}$ )의  $S \times 1$  열 벡터를 나타내고,  $h$ 는 최소 학습 샘플의 인덱스 벡터를 나타내며,  $m(m \in \{0, \dots, S-1\})$ 은 각 픽셀( $i$ )에 대응하는 스칼라 인덱스(scalar index)를 나타낸다. 그리고  $I$ 는 항등 행렬,  $L$ 은 수학식 7에 의해 정의되는 공간 변이 라플라시안 행렬을 나타낸다.

### 수학식 7

$$L(m, n) = \begin{cases} \sum_{k \in \mathcal{N}_4(m)} w_{m,k}(I) & n = m \\ -w_{m,n}(I) & n \in \mathcal{N}_4(m) \\ 0 & \text{otherwise} \end{cases}$$

여기서  $m$  및  $n$ 은 픽셀에 대응하는 스칼라 인덱스를 나타내고,  $\mathcal{N}_4(m)$ 은 픽셀  $m$ 에 대한 인접 픽셀의 집합을 나타낸다.

보간 맵 생성부(153)는 획득된 보간 맵( $p$ )을 양성 샘플 추출부(151)로 전달하고, 양성 샘플 추출부(151)는 보간 맵 생성부(153)에서 전달된 보간 맵( $p$ )에 대해 다시 대응점 일관성을 이용하여 양성 샘플을 학습 샘플( $\hat{p}_i$ )로서 획득한다. 이때 양성 샘플 추출부(151)는 학습 샘플( $\hat{p}_i$ )의 공간적 위치( $\Omega$ )를 함께 획득하여 학습 맵 획득부(155)로 전달한다.

학습 맵 획득부(155)는 양성 샘플 추출부(151)에서 전달되는 학습 샘플( $\hat{p}_i$ )과 위치( $\Omega$ )에 따라 학습 맵( $\hat{p}$ )을 획득하고, 획득된 학습 맵( $\hat{p}$ )을 디스패리티 맵 획득부(140)로 전달한다.

- [0071] 이에 디스패리티 맵 획득부(140)는 현재 인코더(120)의 삼 CNN에서 획득한 특징 맵( $A^l$ ,  $A^r$ )으로부터 획득된 디스패리티 맵( $d$ )과 학습 맵( $\hat{p}$ ) 사이의 오차를 매칭 비용 계산부(130)를 통해 역전파함으로써, 삼 CNN의 가중치( $w$ )가 학습 레이블(또는 학습 레이블이 포함된 학습 영상)을 기반으로 추정되는 오차에 따라 업데이트 되도록 한다. 즉 삼 CNN을 학습시킨다. 이러한 학습은 오차가 기지정된 기준 오차 이내로 감소될 때까지 반복적으로 수행될 수 있으며, 다수의 입력 스테레오 영상에 대해 각각 수행된다.
- [0072] 도3 에서 학습 맵 생성부(150)의 상세 동작을 살펴보면, 우선 양성 샘플 추출부(151)가 디스패리티 맵( $d$ )에서 최소 양성 샘플( $\hat{d}_i$ )을 추출하고, 이후 보간 맵 생성부(153)가 보간 맵( $p$ )을 생성하여 다시 양성 샘플 추출부(151)로 전달한다. 이에 양성 샘플 추출부(151)는 보간 맵( $p$ )으로부터 학습 샘플( $\hat{p}_i$ )과 위치( $\Omega$ )를 추출하여 학습 맵 획득부(155)로 전달하고, 학습 맵 획득부(155)는 학습 샘플( $\hat{p}_i$ )과 위치( $\Omega$ )를 이용하여 학습 맵( $\hat{p}$ )을 획득한다.
- [0073] 상기한 바와 같이 학습 맵 생성부(150)가 입력된 스테레오 영상으로부터 에피폴라 제약에 따른 대응점 일관성을 이용하여 양성 샘플을 추출하고, 추출된 양성 샘플을 전파하여 학습 레이블을 획득하는 경우, 학습 레이블이 포함된 별도의 학습 영상을 획득할 필요가 없다는 장점이 있다.
- [0074] 또한 야외와 같이 조명의 변화가 강하게 나타나는 환경에서 획득된 스테레오 영상에 대해서도 신뢰성 있게 스테레오 매칭을 수행하도록 학습시킬 수 있으며, 폐색이나 질감없는 영역의 모호함에도 불구하고 높은 스테레오 매칭 성능을 나타내도록 학습시킬 수 있다.
- [0075] 도3 에서 우측의 이미지( $d^l$ ,  $\hat{d}^l$ ,  $\hat{p}^l$ )은 각각 좌 디스패리티 맵, 좌 최소 양성 샘플 맵 및 좌 학습 맵을 나타낸다. 도3 도시된 바와 같이, 초기 디스패리티 맵( $d^l$ )에 비해 최소 양성 샘플 맵( $\hat{d}^l$ )에서 더욱 정확하게 객체가 표현됨을 알 수 있다. 그리고 학습 맵( $\hat{p}^l$ )은 최소 양성 샘플 맵( $\hat{d}^l$ )에서 누락된 픽셀이 보완됨으로써 영상 품질이 개선되었음을 알 수 있다.
- [0076] 여기서는 간단한 예시로서 좌 디스패리티 맵( $d^l$ ), 좌 최소 양성 샘플 맵( $\hat{d}^l$ ) 및 좌 학습 맵( $\hat{p}^l$ )을 도시하였으나, 학습 맵 생성부(150)에서는 우 디스패리티 맵( $d^r$ ), 우 최소 양성 샘플 맵( $\hat{d}^r$ ) 및 우 학습 맵( $\hat{p}^r$ ) 또한 유사하게 획득된다.
- [0077] 도4 는 본 발명의 실시예에 따라 생성된 학습 맵의 예를 나타낸다.
- [0078] 도4 에서 (a)는 입력 스테레오 영상의 좌 영상을 나타내고, (b)는 본 실시예에 따라 생성된 학습 맵을 나타내고, (c)는 KITTI 벤치 마크에서 제공된 검증 자료(ground truth)를 이용하여 생성된 디스패리티 맵을 나타낸다.
- [0079] 도4 에 도시된 바와 같이, 본 실시예에 따라 검증 자료가 포함되지 않은 스테레오 영상에서 획득된 학습 맵이 검증 자료가 포함된 디스패리티 맵에 비해 매우 양호한 수준의 학습용 디스패리티 맵을 제공할 수 있다.
- [0080] 한편, 도3 에서 스테레오 영상의 좌 영상 및 우 영상 중 대응하는 영상을 각각 인가받아 특징 맵( $A^l$ ,  $A^r$ )을 생성하는 삼 CNN의 구조는 다양하게 설계될 수 있다.
- [0081] 본 실시예에서는 일례로 삼 CNN의 구조를 심플 CNN 구조와 정밀 CNN 구조로 2가지 구조를 제안한다.
- [0082] 우선 심플 CNN 구조가 표1 에 나타나 있다. 심플 CNN은 빠른 동작이 3 X 3 크기의 컨볼루션 커널(convolution kernel)로 구성되는 4개의 컨볼루션 레이어(convolution layer)(conv1 ~ conv4)를 포함한다. 그리고 3개의 컨볼루션 레이어(conv1 ~ conv3)에는 활성화 함수(activation function)로서 ReLU(Rectifier Linear Unit)가 적용된다. ReLU는 각 컨볼루션 레이어(conv1 ~ conv3)의 인코딩 결과에서 양수 부분만 다음 레이어로 전달되도록 한다.
- [0083] 그리고 마지막 컨볼루션 레이어(conv4)에는 ReLU 대신 L2 정규화 레이어( $l_2$ -norm)를 추가하여 음수의 인코딩 결

과가 유지되도록 한다.

[0084] 표1 에서 Correlation은 도1 의 매칭 비용 계산부(130)의 구성으로, 매칭 비용 계산부(130)는 매칭 비용 볼륨(C)을 획득하기 위해 입력되는 스테레오 이미지의 크기가  $h \times w$ 일때, 디스패리티 후보( $d=\{1, \dots, d_{\max}\}$ )에 대해  $h \times w \times d_{\max}$  횟수의 연산을 필요로 한다.

표 1

Layer	Kernel	Channels	In	Out	Input
conv1	$3 \times 3$	3/64	1	1	Stereo images
conv2	$3 \times 3$	64/64	1	1	conv1
conv3	$3 \times 3$	64/64	1	1	conv2
conv4	$3 \times 3$	64/64	1	1	conv3
$L_2$ norm	$1 \times 1$	64/64	1	1	conv4
Correlation	$1 \times 1$	$64/d_{\max}$	1	1	$L_2$ norm

[0085]

[0086] 표1 의 심플 CNN으로도 일정 수준 이상의 성능을 나타낼 수 있으나, 각 컨볼루션 레이어(conv1 ~ conv4)에서 출력되는 특징 맵의 크기를 조절하지 않는 단일 스케일(single scale) 구조로 식별력(discriminative power)과 수용 영역(receptive field)에 한계가 있다.

[0087] 표2 에 나타난 정밀 CNN 구조에서는 서브 샘플링 레이어와 스킵 커넥션을 포함하는 U-net 구조를 이용한다. 이는 피드 포워드 프로스를 통해 다중 스케일(multi scale) 정보를 증가시킨다.

[0088] 정밀 CNN 구조에서는 수렴 경로(contracting path)와 확장 경로(expansive path)가 포함되며, 수렴 경로에서는 일반적인 CNN과 마찬가지로 2개의  $3 \times 3$  컨볼루션 커널을 연속하여 적용하고, 각각 배치 정규화(batch normalization) 및 ReLU를 수행한 후, 다운 샘플링을 위해 맥스 풀링(max pooling)을 수행할 수 있다. 이때 스트라이드(stride)는 2로 설정될 수 있다. 각 다운 샘플링 단계에서 채널의 수는 2배로 증가된다.

[0089] 반면 확장 경로에서는 각 단계별로 특징 맵의 업 샘플링과 업 컨볼루션 및 채널의 수를 반으로 줄이고, 2개의  $3 \times 3$  컨볼루션 커널과 배치 정규화 및 ReLU가 적용된다. 그리고 마지막 레이어에서 배치 정규화 및 ReLU 대신  $L_2$  정규화 레이어( $l_2$ -norm)를 추가하여 음수의 인코딩 결과가 유지되도록 한다.

[0090] 표2 에서도 Correlation은 도1 의 매칭 비용 계산부(130)의 구성으로, 매칭 비용 볼륨(C)을 계산한다.

표 2

Layer	Kernel	Channels	In	Out	Input
conv1	$3 \times 3$	3/64	1	1	Stereo images
conv1b	$3 \times 3$	64/64	1	1	conv1
pool1	$2 \times 2$	64/64	1	2	conv1b
conv2	$3 \times 3$	64/128	2	2	pool1
conv2b	$3 \times 3$	128/128	2	2	conv2
pool2	$2 \times 2$	128/128	2	4	conv2b
conv3	$3 \times 3$	128/256	4	4	pool2
conv3b	$3 \times 3$	256/256	4	4	conv3
pool3	$2 \times 2$	256/256	4	8	conv3b
conv4	$3 \times 3$	256/512	8	8	pool3
conv4b	$3 \times 3$	512/512	8	8	conv4
pool4	$2 \times 2$	512/512	8	16	conv4b
conv5	$3 \times 3$	512/1024	16	16	pool4
conv5b	$3 \times 3$	1024/512	16	16	conv5
upconv4	$4 \times 4$	512/512	16	8	conv5b
conv6	$3 \times 3$	1024/512	8	8	upconv4,conv4b
conv6b	$3 \times 3$	512/256	8	8	conv6
upconv3	$4 \times 4$	256/256	8	4	conv6b
conv7	$3 \times 3$	512/256	4	4	upconv3,conv3b
conv7b	$3 \times 3$	256/128	4	4	conv7
upconv2	$4 \times 4$	128/128	4	2	conv7b
conv8	$3 \times 3$	256/128	2	2	upconv2,conv2b
conv8b	$3 \times 3$	128/64	2	2	conv8
upconv1	$4 \times 4$	64/64	2	1	conv8b
conv9	$3 \times 3$	128/64	1	1	upconv1,conv1b
conv9b	$3 \times 3$	64/64	1	1	conv9
$L_2$ norm	$1 \times 1$	64/64	1	1	conv9b
Correlation	$1 \times 1$	$64/d_{\max}$	1	1	$L_2$ norm

[0091]

[0092]

상기에서는 삼 CNN의 구조의 일례로 심플 CNN 구조와 정밀 CNN 구조로 2가지 구조를 제안하였으나, 본 발명에는 한정되지 않는다.

[0093]

도5 는 본 발명의 일 실시예에 따른 스테레오 매칭 방법 및 이의 학습 방법에 대한 개략적 구성을 나타내고, 도 6 은 도5 의 학습 맵 생성 및 학습 단계를 상세하게 나타내며, 도7 은 도4 의 학습 맵 생성 및 학습 단계의 알고리즘을 나타낸다.

[0094]

도5 내지 도7 을 참조하여 본 실시예에 따른 스테레오 매칭 방법 및 이의 학습 방법을 설명하면, 우선 스테레오 매칭 장치를 학습 시키기 위해, 다수의 학습 스테레오 영상이 스테레오 영상 입력부(110)로 입력된다(S10). 여기서 다수의 학습 스테레오 영상은 단순히 학습을 위해 이용되는 스테레오 영상을 의미하는 것으로, 학습 레이블이 포함되지 않은 일반의 스테레오 영상이다. 즉 본 실시예에서는 스테레오 매칭 장치의 학습을 위해 학습 레이블이 포함된 별도로 제작된 학습용 스테레오 영상을 요구하지 않는다.

[0095]

그리고 다수의 학습 스테레오 영상 각각에 대해 스테레오 매칭 방식과 동일한 방식으로 학습 디스패리티 맵(d)을 획득한 후, 획득된 디스패리티 맵(d)로부터 학습 맵을 생성하고 학습한다(S20).

[0096]

도6 및 도7 을 참조하여, 학습 맵 생성 및 학습 단계를 상세하게 설명하면, 인코더(120)에서 학습이 완료되지 않은 동일한 구조와 동일한 가중치(w)를 갖는 2개의 삼 CNN이 학습 스테레오 영상으로부터 학습 특징 맵( $A^l$ ,  $A^r$ )을 추출한다(S21). 그리고 매칭 비용 계산부(130)는 특징 맵( $A^l$ ,  $A^r$ )으로부터 학습 매칭 비용 볼륨( $C^l$ ,  $C^r$ )을 계산한다(S22).

[0097]

디스패리티 맵 획득부(140)는 획득된 매칭 비용 볼륨( $C^l$ ,  $C^r$ )의 모든 픽셀(i)에 대해 디스패리티 후보들( $d=\{1,$



$\dots, d_{\max}\}$ ) 중 매칭 비용 볼륨( $C^l, C^r$ )이 최소화되는 디스패리티(d)를 수학적 식 3에 따라 획득하여 학습 디스패리티 맵(d)을 생성한다(S23).

[0098] 학습 특징 맵 추출 단계(S21)부터 디스패리티 맵 생성 단계(S23)까지는 스테레오 매칭 장치가 스테레오 매칭을 수행하는 방법과 동일하다. 다만 상기한 바와 같이 학습 시에는 인코더(120)의 삼 CNN이 학습 완료되지 않은 상태라는 스테레오 매칭 동작과 상이하다.

[0099] 학습 디스패리티 맵(d)이 획득되면, 학습 맵 생성부(150)의 양성 샘플 추출부(151)가 학습 디스패리티 맵(d)에 대해 에피폴라 제약에 따른 대응점 일관성에 기반하여 최소 양성 샘플( $\hat{d}_i$ )을 추정한다(S24). 이때 양성 샘플 추출부(151)는 추정된 최소 양성 샘플( $\hat{d}_i$ ) 이외의 나머지 픽셀은 제거한다.

[0100] 한편, 보간 맵 생성부(153)는 입력 컬러 영상을 가이드로 이용하여 최소 양성 샘플( $\hat{d}_i$ )을 인접 픽셀로 전파하는 보간 기법에 따라 보간 픽셀(pi)을 획득함으로써, 보간 맵(p)을 생성한다(S25).

[0101] 양성 샘플 추출부(151)는 보간 맵(p)에 대해 다시 대응점 일관성에 기반하여 학습 샘플( $\hat{p}_i$ )의 공간적 위치( $\Omega$ )를 추정한다(S26). 그리고 학습 맵 획득부(155)는 양성 샘플 추출부(151)에서 전달되는 학습 샘플( $\hat{p}_i$ )과 위치( $\Omega$ )에 따라 학습 맵( $\hat{p}$ )을 획득한다(S27).

[0102] 획득된 학습 맵( $\hat{p}$ )은 디스패리티 맵 획득부(140)로 전달되고, 디스패리티 맵 획득부(140)는 디스패리티 맵(d)과 학습 맵( $\hat{p}$ ) 사이의 오차를 매칭 비용 계산부(130)를 통해 역전파함으로써, 삼 CNN의 가중치(w)가 업데이트 되도록 한다(S28). 즉 삼 CNN을 학습시킨다.

[0103] 상기한 학습은 오차가 기지정된 기준 오차 이내로 감소될 때까지 반복적으로 수행될 수 있으며, 다수의 입력 스테레오 영상에 대해 각각 반복적으로 수행된다.

[0104] 다시 도5 를 참조하면, 인코더(120)의 삼 CNN에 대한 학습이 완료된 이후, 스테레오 매칭을 수행할 스테레오 영상이 스테레오 영상 입력부(110)로 입력된다.

[0105] 그리고 인코더(120)에서 학습이 완료된 2개의 삼 CNN이 특징 맵( $A^l, A^r$ )을 추출한다(S40). 매칭 비용 계산부(130)는 추출된 특징 맵( $A^l, A^r$ )으로부터 매칭 비용 볼륨( $C^l, C^r$ )을 계산한다(S50).

[0106] 디스패리티 맵 획득부(140)는 매칭 비용 볼륨( $C^l, C^r$ )이 최소화되는 디스패리티(d)를 수학적 식 3에 따라 획득하여 디스패리티 맵(d)을 생성함으로써, 스테레오 매칭을 수행한다(S60).

[0107] 한편, 본 발명의 실시예에 따른 학습 맵 생성부(150)는 대응점 일관성을 이용하여 불일치 픽셀을 제거하며 오차 역전파 과정이 포함됨으로써, 손실이 발생할 수 있다. 이는 모든 가능한 디스패리티 후보들에 대한 각 픽셀의 소프트 맥스 손실(softmax loss)을 허용한다.

[0108] 본 실시예에서는 각 픽셀 및 디스패리티 후보들에 대해 엔트로피 손실을 계산하며, 엔트로피 손실은 수학적 식 8로 정의된다.

### 수학적 식 8

$$\mathcal{L}_w = -\frac{1}{N} \sum_{i \in \Omega} \sum_s P_T(s; i) \log(P(s; i))$$

[0110] 여기서 s는  $s_x = \{i_x - 1, \dots, i_x - d_{\max}\}$ 로 모든 디스패리티에 대해 정의되는 값이며,  $P_T(s; i)$ 는 샘플링 된 학습 집합에 대해 1이고, 이외에는 0으로 정의되는 레이블이고,  $P(s; i)$ 는 수학적 식 9와 같이 정의되는 소프트 맥스

확률이다.

### 수학식 9

$$P(s; i) = \frac{\exp(-\|\mathcal{F}_{\mathbf{W}}(I^l(i)) - \mathcal{F}_{\mathbf{W}}(I^r(s))\|_1)}{\sum_v \exp(-\|\mathcal{F}_{\mathbf{W}}(I^l(i)) - \mathcal{F}_{\mathbf{W}}(I^r(v))\|_1)}$$

여기서  $v$ 는  $s$ 와 마찬가지로,  $v_x = \{i_x - 1, \dots, i_x - d_{\max}\}$ 로 모든 디스패리티에 대해 정의되는 값이다. 수학식 9에서는 음의 부호를 적용함으로써, 유사성 스코어를 매칭 비용으로 변환한다.

학습된 CNN을 이용하더라도, 원본 매칭 비용 볼륨으로는 정확한 디스패리티 맵을 생성하기에 충분하지 않을 수 있다. 특히 질감 없는 영역과 폐색 영역에서 오류가 발생할 수 있다. 이에 기존의 스테레오 매칭 기법과 마찬가지로 후처리 작업이 더 추가될 수 있다.

후처리 작업은 글로벌 매칭과 이후 일관성 검사, 서브 픽셀 강조, 중간값 및 양방향 필터링 작업이 포함될 수 있다.

반 전역 매칭(Semi-Global Matching)은 디스패리티 맵에 대해 평활 제약(smoothness constraint) 조건을 적용하여 매칭 비용을 저감 시킨다.

디스패리티 맵( $d$ )의 에너지 함수( $E(d)$ )를 수학식 10과 같이 정의할 수 있다.

### 수학식 10

$$E(d) = \sum_i \{C(i, d(i)) + \sum_{j \in N(i)} P_1 [|d(i) - d(j)| = 1] + \sum_{j \in N(i)} P_2 [|d(i) - d(j)| > 1]\},$$

수학식 10에서  $\sum_i \{C(i, d(i)) + \sum_{j \in N(i)} P_1 [|d(i) - d(j)| = 1] + \sum_{j \in N(i)} P_2 [|d(i) - d(j)| > 1]\}$ 는 매칭 비용이 높은 디스패리티에 패널티를 부가하고,  $\sum_{j \in N(i)} P_1 [|d(i) - d(j)| = 1]$  및  $\sum_{j \in N(i)} P_2 [|d(i) - d(j)| > 1]$ 는 각각 불연속적인 디스패리티와  $P_1 < P_2$ 의 조건에 대해 패널티를 부가한다.

반 전역 매칭 기법에서는 여러 방향에 따라 동적 프로그래밍 스타일로 1차원 비용 업데이트를 수행하여 근사값을 획득하며, 일반적으로 두개의 수평 방향과 2개의 수직 방향에 따라 최적화 한다. 이에 방향  $r$ 에서의 수학식 10의 에너지 함수( $E(d)$ )를 최소화하기 위해 방향  $r$ 을 따른 매칭 비용( $L_r(i, d(i))$ )을 수학식 11과 같이 정의한다.

### 수학식 11

$$L_r(i, d) = C(i, d) + \min\{L_r(i - \mathbf{r}, d), L_r(i - \mathbf{r}, d \pm 1) + P_1 \min_k (L_r(i - \mathbf{r}, k) + P_2)\}.$$

수학식 11을 이용하여 4개의 방향에 대한 비용을 평균한 후, 좌측 및 우측 방향에 대한 초기 디스패리티 맵을 계산할 수 있다.

이후 보간, 서브 픽셀 강조 및 순화와 같은 후처리를 통해 초기 디스패리티 맵을 개선할 수 있다. 보간은 좌우 일관성 검사를 수행하여 좌 디스패리티 맵과 우 디스패리티 맵 사이의 충돌을 해소하며, 서브 픽셀 강조는 이차 함수를 인접한 픽셀에 적용하여 서브 픽셀이 강조된 디스패리티 맵을 획득한다.



- [0123] 이후, 에지를 흐리게 하지 않고 디스패리티 맵을 평활화하기 위해, 5 x 5 중간값 필터 및 양측 필터를 순차적으로 적용할 수 있다.
- [0124] 이하에서는 본 실시예에 따른 스테레오 매칭 장치 및 방법의 성능을 시뮬레이션한 결과를 나타낸다.
- [0125] VLFeat MatConNet Toolbox를 사용하여 시뮬레이션 환경을 구성하였으며, 학습을 위한 스테레오 영상은 알려진 KITTI, Middlebury, HCI 및 Yonsei 벤치 마크 스테레오 영상에서 임의의 256 X 512 크기로 잘라낸 패치를 이용하였으며, 최대 디스패리티 범위( $d_{max}$ )는 228, 116, 140, 80으로 설정되었다.
- [0126] KITTI 2012와 KITTI 2015에서 각각 194와 200 쌍의 스테레오 영상과 2005년과 2006년 사이의 Middlebury의 27 쌍의 스테레오 영상을 비지도 방식으로 학습하였으며, 특히 조명의 변화에 대한 신뢰성을 평가하기 위해, Middlebury에서 3가지 다른 조건을 갖는 스테레오 영상을 이용하였다. 또한, 다양한 야외 운전 조건을 평가하기 위해, 11개의 도전 시나리오가 포함된 HCI 벤치 마크와 시간 및 계절의 조건이 상이한 53개의 Yonsei 벤치 마크 스테레오 영상을 이용하였다.
- [0127] 그리고 임계값( $t$ )은 1, 밸런스 상수( $\lambda$ )는  $20^2$ , 범위 파라미터( $\sigma_c$ )는 0.015로 설정되었다. 또한 학습 최적화를 위해 초기 학습율(initial learning rate)과 모멘텀(momentum) 및 가중치 감쇄(weight decay)는 각각 0.001, 0.9 및 0.005로 설정되었다.
- [0128] 표3 은 본 실시예에 따라 생성된 학습 맵( $\hat{P}$ )을 이용한 비지도 학습 방법을 다른 비지도 학습 방법과 비교하여 시뮬레이션 한 결과로 MIDDLEBURY, KITTI 2012, 및 KITTI 2015 학습 데이터에 대한 비지도 학습 방법에서의 오차율의 정량적 비교 결과를 나타낸다.

### 표 3

Method	Error rates(%)		
	Middlebury	KITTI 2012	KITTI 2015
Col. Consistency	34.47	17.08	17.13
Col. & Disp. Consistency	33.66	16.74	16.35
Ours w/o PP	30.08	15.72	15.51
Ours	<b>28.53</b>	<b>13.92</b>	<b>13.88</b>

- [0129]
- [0130] 표3 에서는 본 실시예에서 대응점 일관성 및 양성 샘플 전파를 수행하는 경우와 수행하지 않는 경우(w/o PP)도 함께 비교하여 시뮬레이션하였으며, 심플 CNN을 이용하였다. 그리고 최종 매칭 비용 볼륨에서 spatial transformer network(STN)의 영상 샘플링과 soft-argmin을 적용하였으며, 색상 일관성(Col. consistency) 또는 색상 및 디스패리티 일관성(Col. Disp. Consistency)를 이용하여 학습 맵을 생성하는 경우를 비교하였다.
- [0131] 표3 에 나타난 바와 같이, 본 실시예에 따른 비지도 학습 방법을 이용하는 경우, 다른 비지도 학습 방법에 비해, 대응점 일관성 및 양성 샘플 전파를 수행하지 않는 경우에도 더 나은 성능을 나타내지만, 수행하는 경우 더 향상된 성능을 나타낼 수 있다.
- [0132] 표4 는 본 실시예의 인코더(120)에서 CNN의 구조에 따른 성능 비교 결과를 나타낸다. 표4 에서는 Census와 MC-CNN fast, Content-CNN, MC-CNN-WS와 표1 및 표2 의 심플 CNN 및 정밀 CNN을 비교한 결과를 나타내며, 표3 에서와 마찬가지로 MIDDLEBURY, KITTI 2012, 및 KITTI 2015 학습 데이터에 대한 비지도 학습 방법에서의 오차율의 정량적 비교 결과를 나타낸다.

표 4

Method	Error rates(%)		
	Middlebury	KITTI 2012	KITTI 2015
Census [6]	64.53	53.52	50.35
MC-CNN fst [4]	29.94	15.44	15.38
Content-CNN [20]	-	9.67	8.71
MC-CNN-WS [41]	29.60	13.90	14.08
Ours_S	28.53	13.92	13.88
Ours_P	<b>24.92</b>	<b>8.34</b>	<b>8.23</b>

[0133]

[0134] 표4 에 나타난 바와 같이 본 실시예에 따른 CNN의 구조를 이용하는 경우, 심플 CNN을 이용하더라도 다른 CNN 구조에 비해 양호한 성능을 나타내며, 정밀 CNN을 이용하는 경우 매우 우수한 성능을 나타낼 수 있다.

[0135] 도8 은 본 발명의 실시예에 따른 학습 맵의 생성 과정에서 생성되는 맵의 예를 나타낸다.

[0136] 도8 에서 (a)는 스테레오 영상 입력부(110)에 입력된 입력 스테레오 영상의 좌 영상을 나타내고, (b)는 디스패리티 맵 획득부(140)에서 WTA 기법에 따라 획득된 좌 영상의 디스패리티 맵을 나타내며, (c)는 학습 맵 생성부(150)의 양성 샘플 추출부(151)에서 대응점 일관성을 이용하여 획득된 최소 양성 샘플 맵( $\hat{d}$ )을 나타낸다.

[0137] 그리고 (d)는 보간 맵 생성부(153)에서 보간된 보간 맵(p)을 나타내고, (e)는 양성 샘플 추출부(151)에서 보간 맵(p)에 대해 다시 대응점 일관성을 이용하여 획득된 학습 맵( $\hat{p}$ )을 나타내며, (f)는 검증 자료 디스패리티 맵을 나타낸다.

[0138] 도9 는 본 발명의 실시예에 따른 학습 샘플과 다른 비지도 학습 방식으로 획득되는 샘플을 비교한 도면이다.

[0139] 도9 에서 (a)는 입력 스테레오 영상의 좌 영상을 나타내고, (b)는 대응점 일관성 및 총 변동 임계값을 적용한 최소 양성 샘플 맵( $\hat{d}$ )을 나타내며, (c)는 MC-CNN 스테레오 알고리즘과 신뢰도 임계값을 적용한 방법을 나타낸다.

[0140] (d) 내지 (h)는 본 실시예에서 초기 랜덤 가중치와 반복 횟수를 1, 3, 5, 7로 설정한 경우에 반복 횟수에 따라 획득되는 학습 맵( $\hat{p}$ )을 나타내고, (i)는 검증 자료 디스패리티 맵을 나타낸다.

[0141] 도8 및 도9 에서 본 실시예에 따라 생성된 학습 맵( $\hat{p}$ )은 검증 자료 디스패리티 맵에 비교하여도 양호한 품질의 학습용 맵을 제공할 수 있음을 알 수 있다. 또한 학습 맵( $\hat{p}$ )을 획득하기 위한 과정을 반복함으로써, 더욱 우수한 학습 맵( $\hat{p}$ )을 획득할 수 있음을 알 수 있다.

[0142] 도10 내지 도12 는 본 발명의 실시예에 따른 CNN의 구조에 따른 학습 샘플을 비교한 도면이다.

[0143] 도10 내지 도12 는 각각 Middlebury, KITTI 2012 및 KITTI 2015 데이터 집합에 대해 디스패리티 맵을 획득한 결과를 나타낸다.

[0144] 도10 내지 도12 각각에서 에서 왼쪽 영상은 입력 스테레오 영상의 좌 영상을 나타내고, 가운데 영상은 표1 의 심플 CNN 구조를 이용하여 획득된 디스패리티 맵을 나타내고, 우측 영상은 표2 의 정밀 CNN 구조를 이용하여 획득된 디스패리티 맵을 나타낸다.

[0145] 도10 및 도12 를 참조하면, 비록 심플 CNN 구조보다 정밀 CNN 구조에서 더욱 정확한 디스패리티 맵이 획득됨을 알 수 있으나, 심플 CNN 구조 및 정밀 CNN 구조 모두에서 학습 맵으로 이용하기에 양호한 디스패리티 맵을 획득

할 수 있음을 알 수 있다.

[0146] 도13 및 도14 는 본 발명의 실시예에 따른 CNN의 구조에 따른 스테레오 매칭 결과를 비교한 도면이다.

[0147] 도13 및 도14 는 각각 KITTI 2012 및 KITTI 2015 데이터 집합에 대해 스테레오 매칭 결과로 획득한 디스패리티 맵을 나타내었으며, 왼쪽 영상은 입력 스테레오 영상의 좌 영상을 나타내고, 가운데 영상은 표1 의 심플 CNN 구조를 이용하여 획득된 디스패리티 맵을 나타내고, 우측 영상은 표2 의 정밀 CNN 구조를 이용하여 획득된 디스패리티 맵을 나타낸다.

[0148] 도13 및 도14 에서도 심플 CNN 구조보다 정밀 CNN 구조에서 더욱 정확한 디스패리티 맵이 획득됨을 알 수 있으나, 심플 CNN 구조 및 정밀 CNN 구조 모두에서 모두 양호한 스테레오 매칭 결과로서의 디스패리티 맵을 획득할 수 있음을 알 수 있다.

[0149] 표5 는 본 실시예의 인코더(120)에서 CNN의 구조에 따른 스테레오 매칭 성능 비교 결과를 나타낸다. 표4 에서는 기존의 Guided Filter, Census + SGM, DLP, MC-CNN fast, MC-CNN acrt, Content-CNN, MC-CNN-WS와 표1 및 표2 의 심플 CNN 및 정밀 CNN을 비교한 결과를 나타내며, KITTI 2012 및 KITTI 2015 학습 데이터에 대한 비지도 학습 방법에서의 오차율의 정량적 비교 결과를 나타낸다.

표 5

Method	Error rates(%)			
	KITTI 2012		KITTI 2015	
	Noc	All	Noc	All
Guided Filter [8]	11.65	13.76	—	—
Census + SGM [9]	5.76	7.00	5.62	6.38
DLP [42]	5.28	7.21	—	—
MC-CNN-fst [4]	2.82	—	4.65	—
MC-CNN-acrt [4]	<b>2.43</b>	<b>3.63</b>	<b>3.33</b>	<b>3.89</b>
Content-CNN [20]	3.07	4.29	4.00	4.54
MC-CNN-WS [41]	3.02	4.45	4.11	4.97
Ours_S	2.89	4.14	3.92	4.47
Ours_P	2.98	4.22	4.06	4.81

[0150]

[0151] 도15 및 도16 은 본 실시예에 따른 비지도 학습 방법의 조명에 대한 강건성을 시뮬레이션한 결과를 나타낸다.

[0152] 조명의 변화에 따른 효과를 분석하기 위해 모든 영상에 대한 노출 지수를 1로 설정하고, 조명 지수를 1에서 3으로 변경하였다. 즉 좌측 및 우측 조명이 1/3인 경우에 대한 시뮬레이션 결과를 나타낸다.

[0153] 그리고 도15 에서는 Middlebury 데이터 집합을 이용하였으며, 왼쪽 2개의 영상은 입력되는 스테레오 영상의 좌 영상 및 우영상을 나타내고, 이후 우측 방향으로 특징 기반의 DASC (Dense Adaptive Self-Correlation)와 색상 및 디스패리티 일관성(Col. Disp. Consistency), 본 실시예 및 검증 자료 디스패리티 맵을 나타낸다.

[0154] 도15 에서 확인할 수 있듯이 본 실시예에 따른 비지도 학습 방법에 따라 학습된 스테레오 매칭 장치는 조명의 변화에도 다른 학습 방식에 비해 매우 우수한 성능을 나타냄을 알 수 있다.

[0155] 그리고 도16 에 도시된 바와 같이 본 실시예에 따른 비지도 학습 방법은 조명의 변화에 대해 심플 CNN을 이용하는 경우에도 다른 학습 방식과 유사한 매칭 비용을 갖지만, 정밀 CNN을 이용하는 경우, 매우 낮은 오차율을 나타냄을 알 수 있다.

[0156] 도17 및 도18 은 야외 운전 환경에서의 스테레오 매칭 성능을 시뮬레이션한 결과를 나타낸다.

[0157] 도17 은 HCI 데이터 집합에 대해 시뮬레이션한 결과를 나타내고, 도18 은 Yonsei 데이터 집합에 대한 시뮬레이션한 결과를 나타낸다.

- [0158] 도17 및 도18 에서 왼쪽 영상은 입력 스테레오 영상의 좌 영상을 나타내고, 가운데 영상은 표1 의 심플 CNN 구조를 이용하여 획득된 디스패리티 맵을 나타내고, 우측 영상은 표2 의 정밀 CNN 구조를 이용하여 획득된 디스패리티 맵을 나타낸다.
- [0159] 도17 및 도18 로부터 본 실시예에 따른 비지도 학습 방법으로 학습된 스테레오 매칭 장치 및 방법은 야외의 실제 운전 환경에서도 우수한 스테레오 매칭 성능을 나타냄을 알 수 있다.
- [0160] 표 6 은 본 실시예에 따른 스테레오 매칭 장치 및 방법의 스테레오 매칭 속도를 시뮬레이션한 결과를 나타낸다.
- [0161] 표6 에서는 DLP 기법과 본 실시예에 따른 심플 CNN 구조 및 정밀 CNN 구조의 인코더를 이용한 경우를 비교한 결과를 나타내며, KITTI 2012, KITTI 2015 및 MIDDLEBURY에 대해 시뮬레이션한 결과를 나타낸다. 또한 KITTI와 MIDDLEBURY에 대해 서로 다른 해상도의 패치를 이용하여 시뮬레이션한 결과를 나타낸다.

표 6

Method	KITTI (1242×375)			Middlebury (656×541)		
	DLP	Ours_S	Ours_P	DLP	Ours_S	Ours_P
Time(ms)	152	193	227	116	155	187

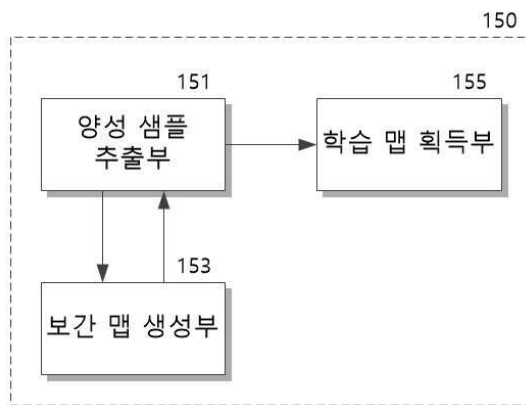
- [0162]
- [0163] 표6 에 나타난 바와 같이 본 실시예에 따른 심플 CNN 구조를 갖는 스테레오 매칭 방법은 DLP에 비해 조금 느리게 동작하지만 스테레오 매칭 성능에 있어서는 상기한 바와 같이 DLP에 비해 월등하다. 그리고 정밀 CNN 구조를 갖는 경우 처리 속도가 느린 반면 매우 우수한 스테레오 매칭 성능을 나타내므로, 효율성과 정확성 사이의 균형을 고려하여 필요한 CNN 구조를 선택적으로 이용할 수 있다.
- [0164] 본 발명에 따른 방법은 컴퓨터에서 실행 시키기 위한 매체에 저장된 컴퓨터 프로그램으로 구현될 수 있다. 여기서 컴퓨터 판독가능 매체는 컴퓨터에 의해 액세스 될 수 있는 임의의 가용 매체일 수 있고, 또한 컴퓨터 저장 매체를 모두 포함할 수 있다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 모두 포함하며, ROM(판독 전용 메모리), RAM(랜덤 액세스 메모리), CD(컴팩트 디스크)-ROM, DVD(디지털 비디오 디스크)-ROM, 자기 테이프, 플로피 디스크, 조명데이터 저장장치 등을 포함할 수 있다.
- [0165] 본 발명은 도면에 도시된 실시예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다.
- [0166] 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 청구범위의 기술적 사상에 의해 정해져야 할 것이다.

도면

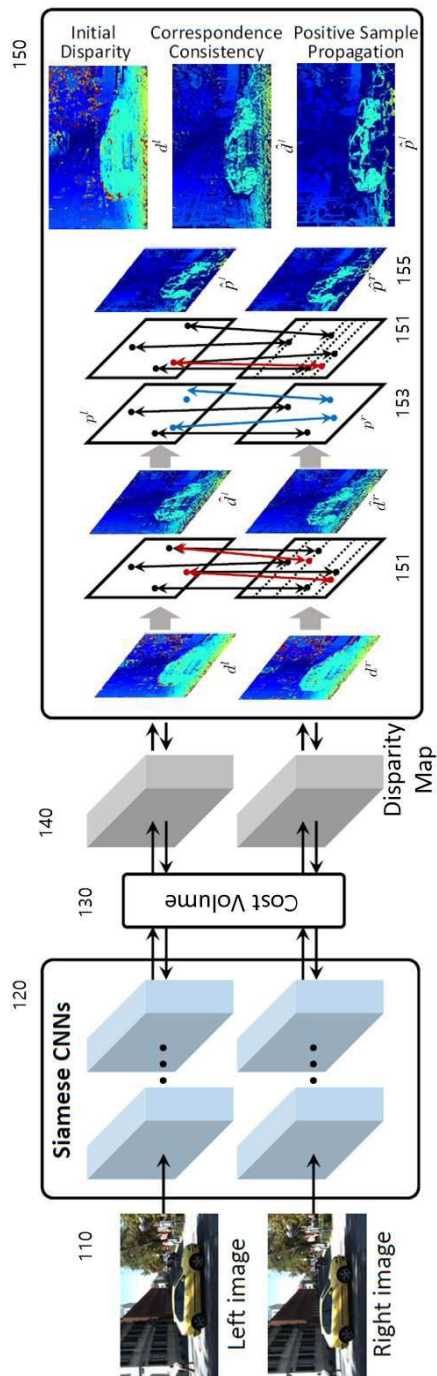
도면1



도면2



도면3

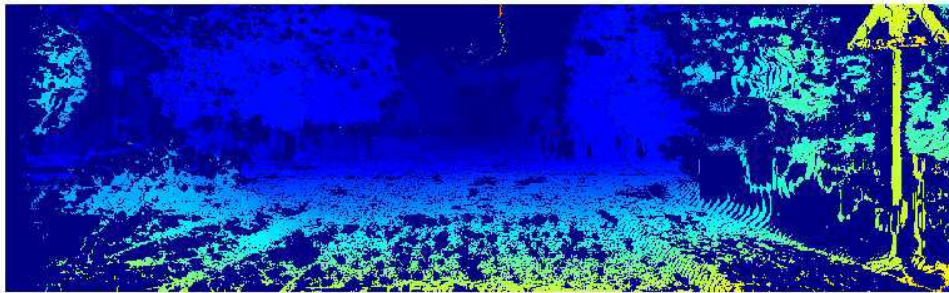




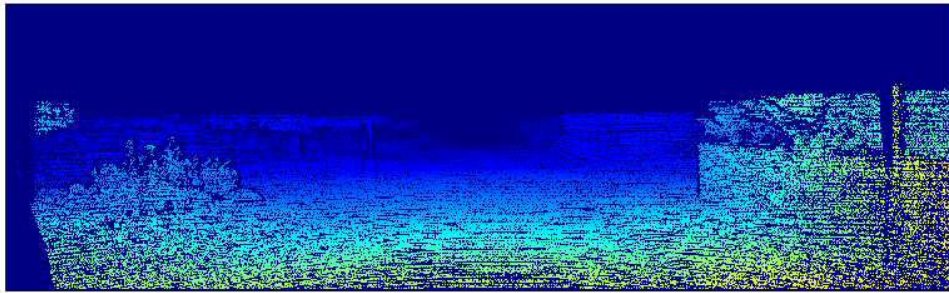
도면4



(a)

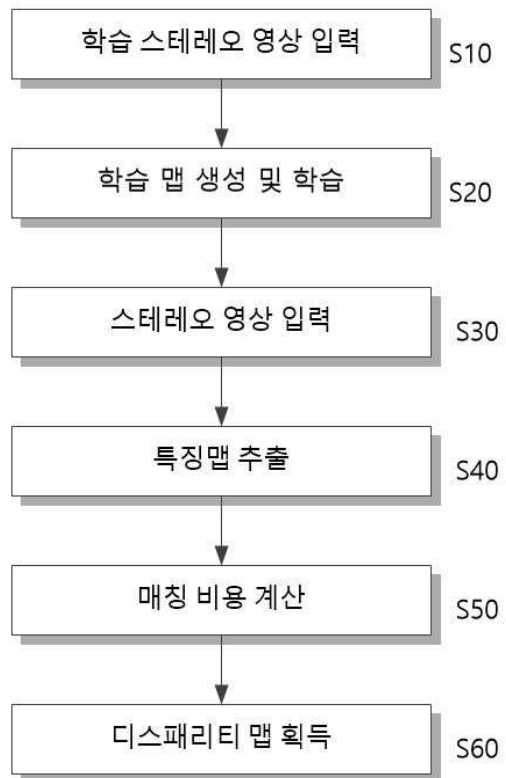


(b)

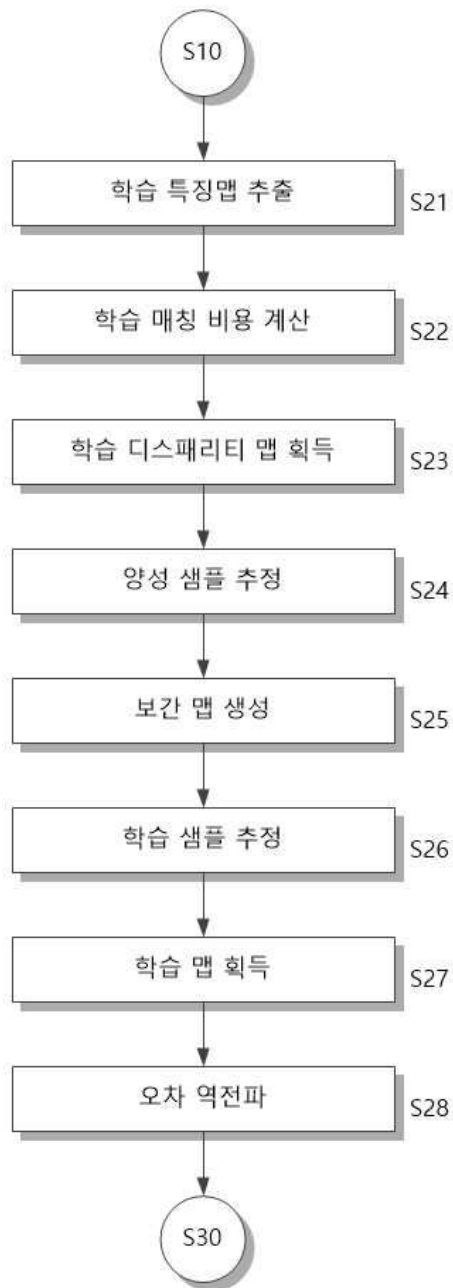


(c)

도면5



도면6



도면7

---

**Algorithm 1:** Unsupervised training scheme of CNNs in each epoch

---

**Input:** Stereo image pairs  $I^l, I^r$ .

**Output:** CNNs parameters  $\mathbf{W}$ .

**Parameters and Notation**

$\mathbf{A} = \mathcal{F}_{\mathbf{W}}(I)$ : Convolutional activation through CNNs with  $\mathbf{W}$ .

$T$ : Total number of stereo image pairs in training sets.

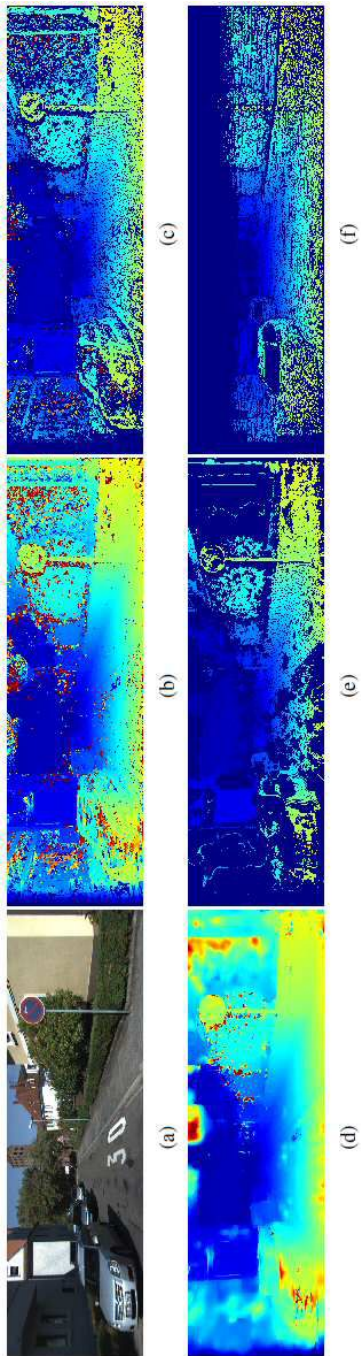
**for**  $t = 1 \dots T$

- 1: Encode an stereo image pair  $I^{l,t}, I^{r,t}$  into convolutional activation feature map  $\mathbf{A}^l, \mathbf{A}^r$ , respectively.
- 2: Construct the matching cost volume  $C^l, C^r$  using  $\mathbf{A}^l, \mathbf{A}^r$  through correlation layer.
- 3: Estimate initial disparity map  $d^l, d^r$  from  $C^l, C^r$  in a WTA strategy.
- 4: Discard inconsistent pixels from  $d^l, d^r$  through correspondence consistency to make sparse disparity map of positive samples  $\hat{d}^l, \hat{d}^r$ .
- 5: Apply a positive sample propagation at  $\hat{d}^l, \hat{d}^r$  to make propagated disparity map  $p^l, p^r$  with  $I^l, I^r$  as guidance.
- 6: Discard inconsistent pixels from  $p^l, p^r$  through correspondence consistency to make sparse disparity map of training samples  $\hat{p}^l, \hat{p}^r$ .
- 7: Update  $\mathbf{W}$  using cross-entropy loss with  $\hat{p}^l, \hat{p}^r$  as training labels.

**end for**

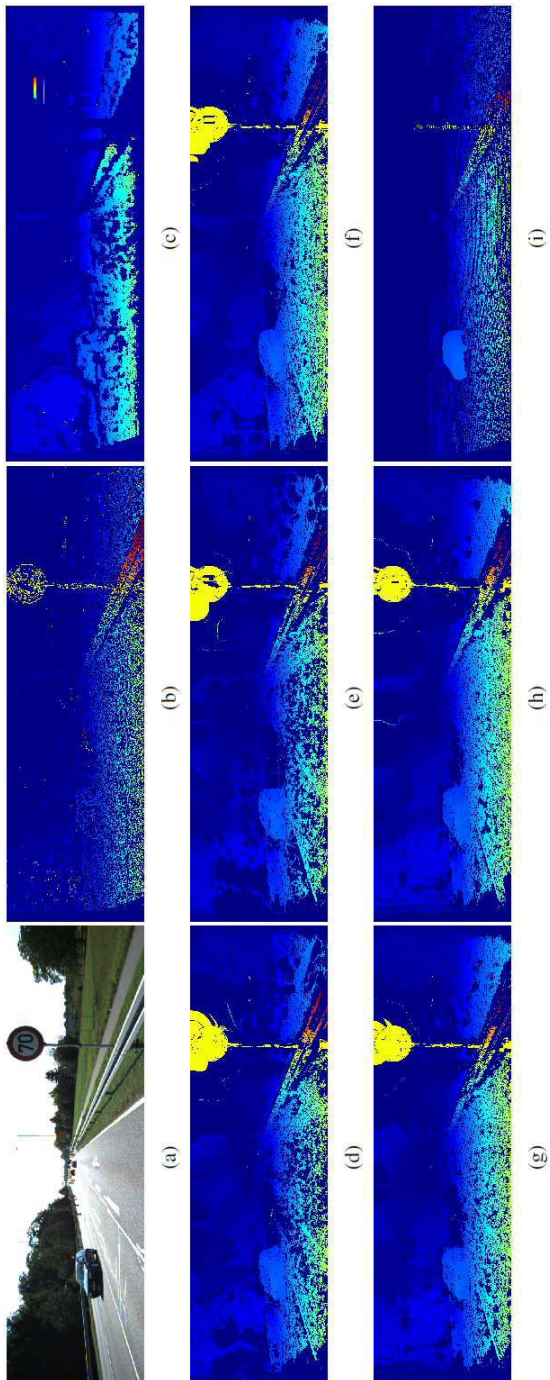
---

도면8

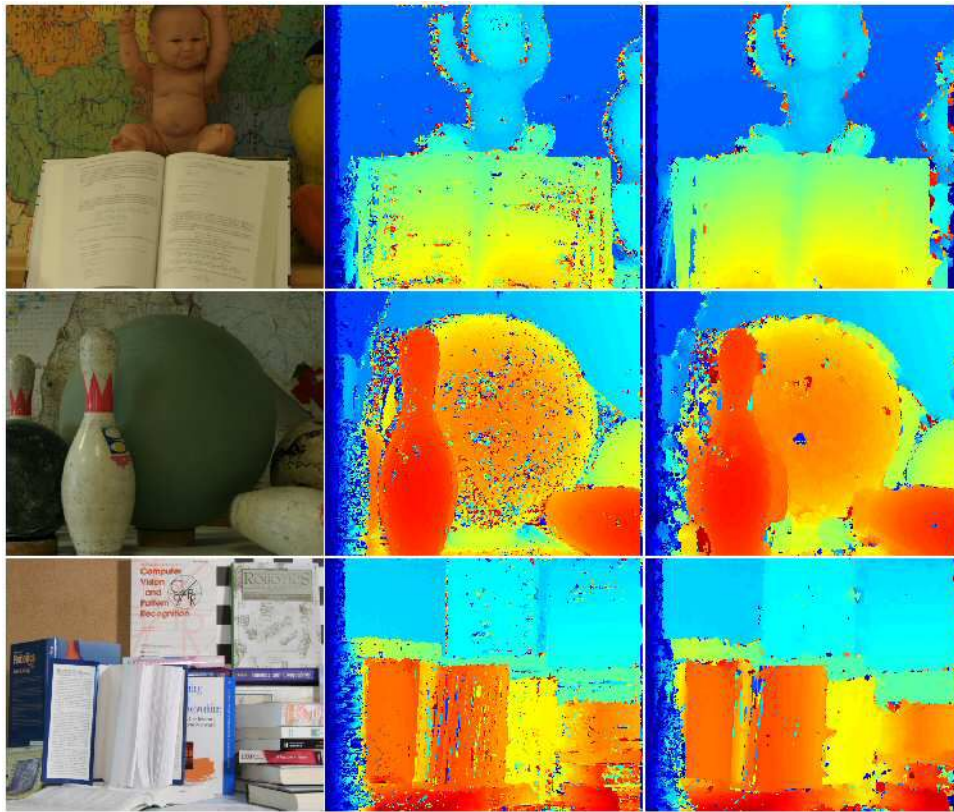




도면9

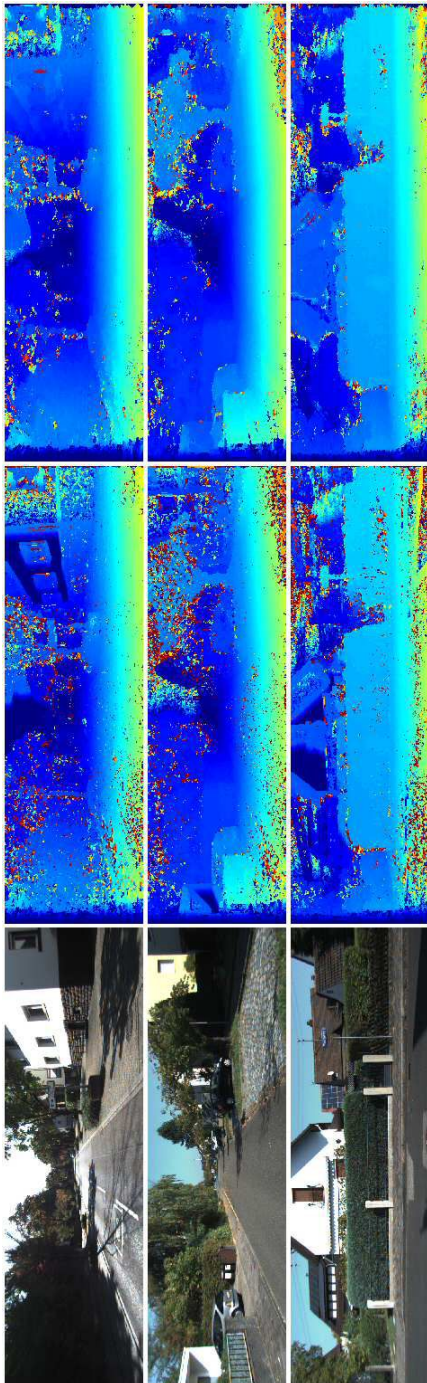


도면10

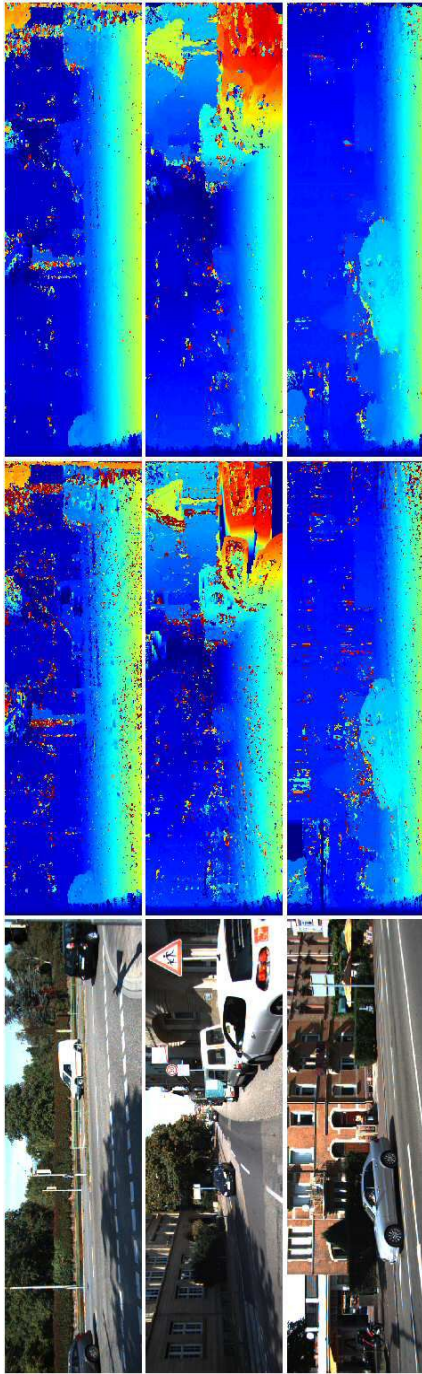




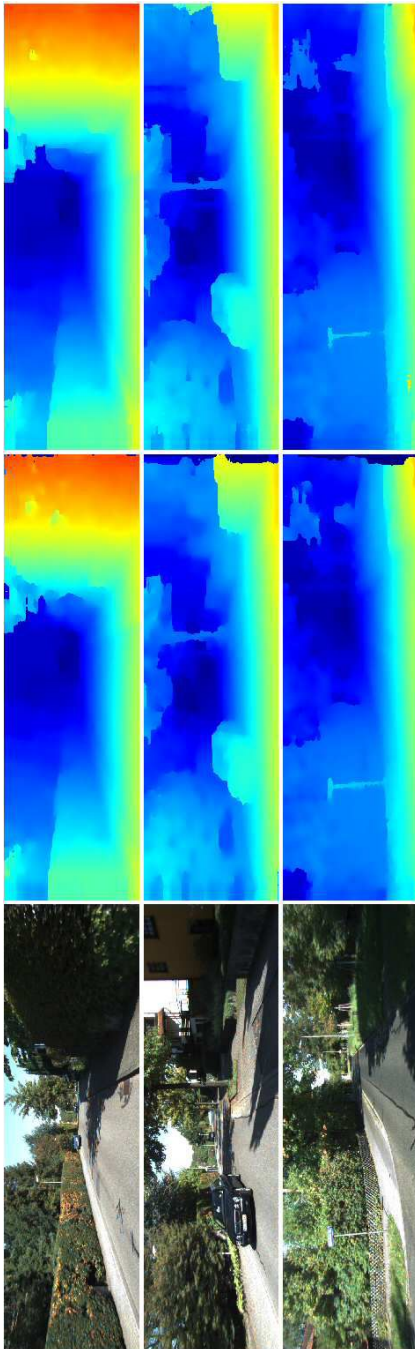
도면11



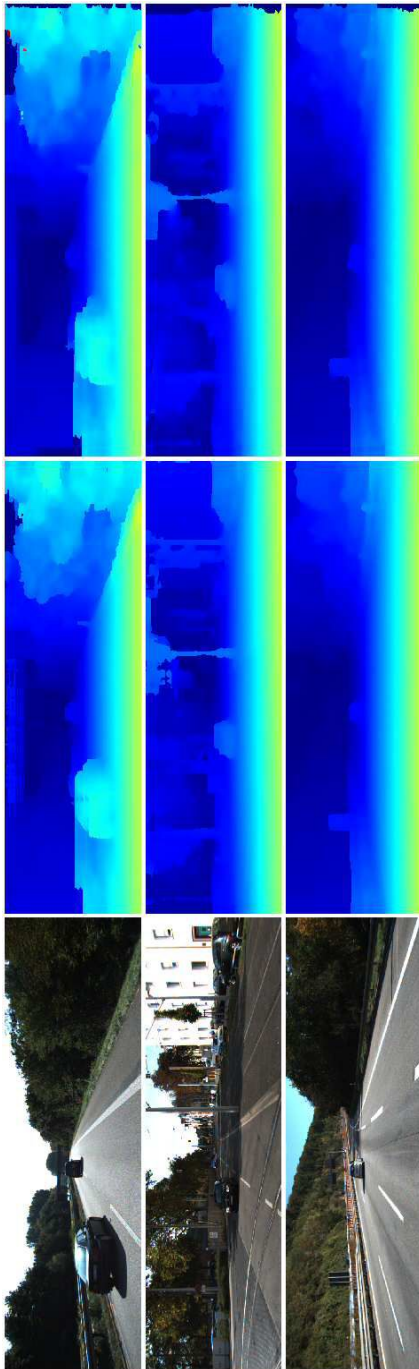
도면12



도면13

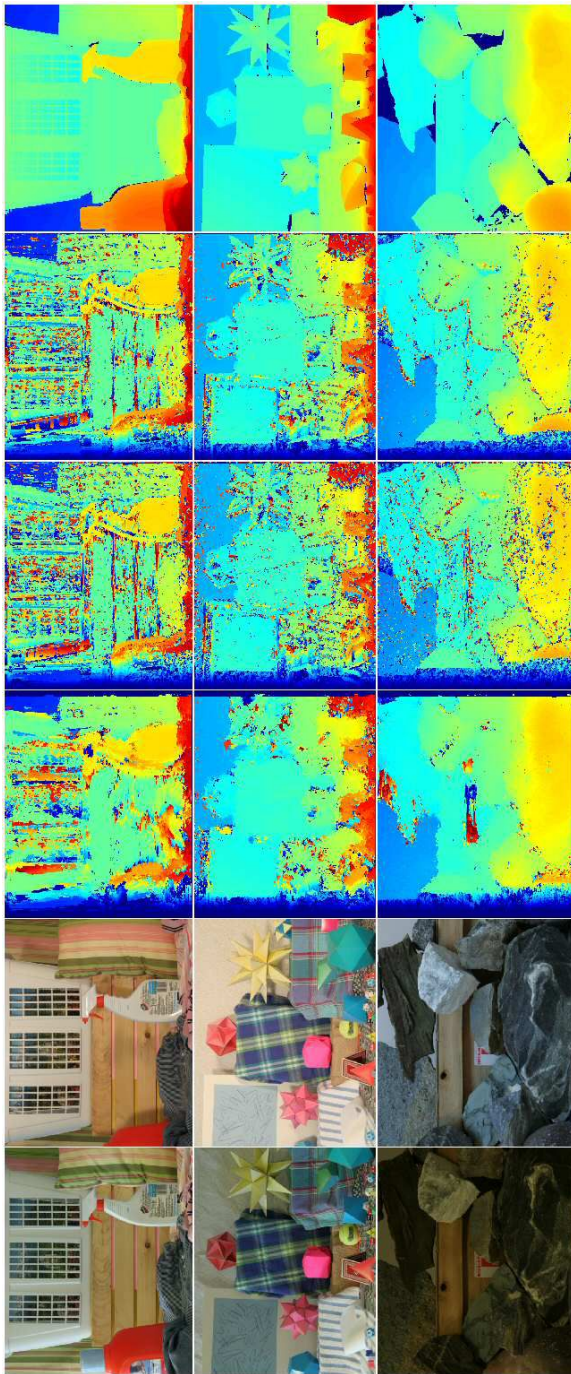


도면14

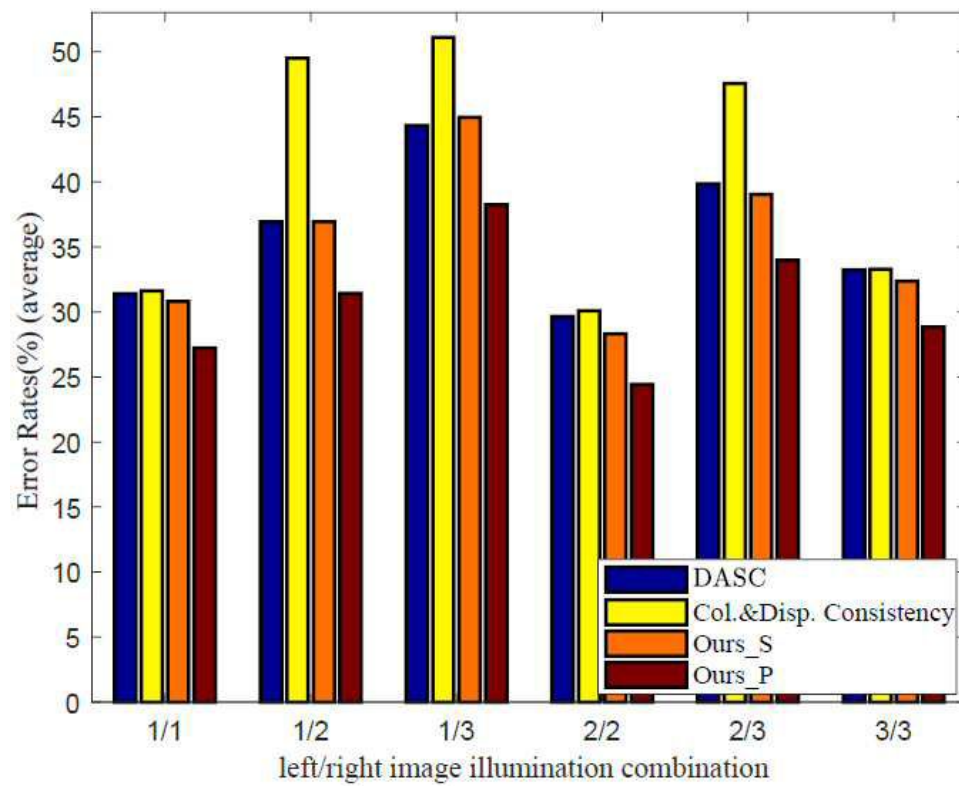




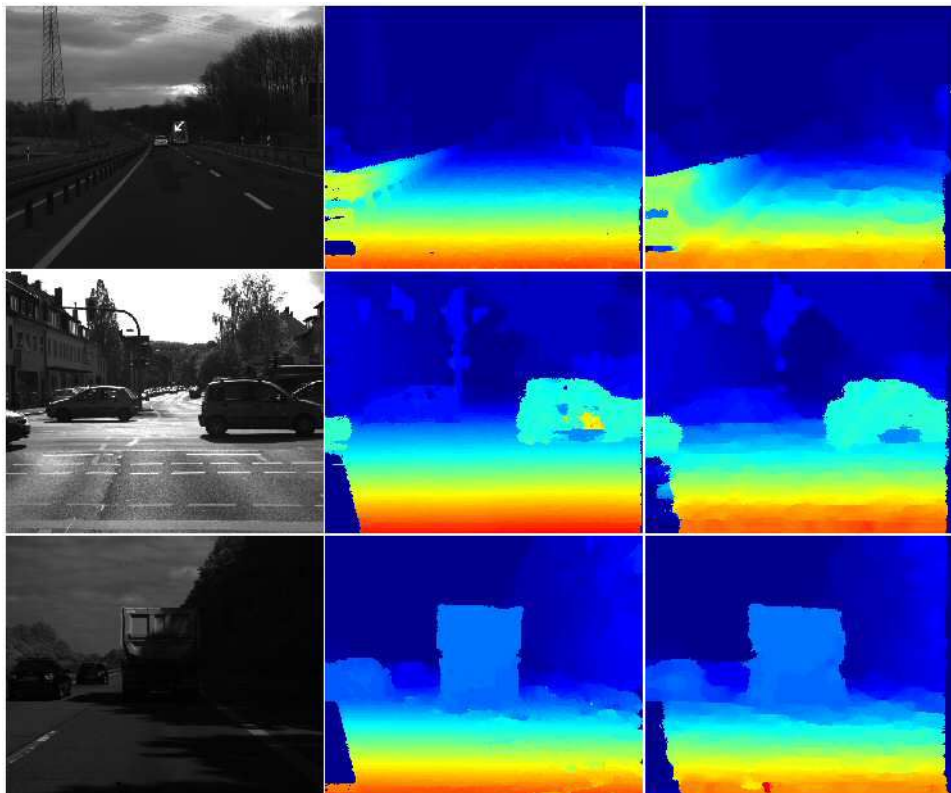
도면15



도면16



도면17



도면18

