



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0095251  
(43) 공개일자 2020년08월10일

(51) 국제특허분류(Int. Cl.)  
G06N 3/08 (2006.01) G06N 3/04 (2006.01)  
H04N 13/204 (2018.01)  
(52) CPC특허분류  
G06N 3/084 (2013.01)  
G06N 3/0454 (2013.01)  
(21) 출원번호 10-2019-0013102  
(22) 출원일자 2019년01월31일  
심사청구일자 2019년01월31일

(71) 출원인  
연세대학교 산학협력단  
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)  
(72) 발명자  
함범섭  
서울특별시 강남구 압구정로61길 37, 72동 506호 (압구정동, 한양아파트)  
박현중  
서울특별시 송파구 양재대로 1218, 204동 504호 (방이동, 올림픽선수기자촌아파트)  
(74) 대리인  
민영준

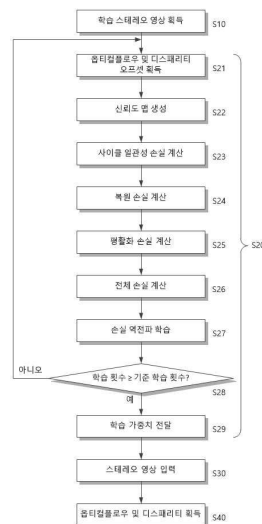
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 사이클 일관성 기반 유틸리티플로우 및 디스패리티 추정 장치 및 방법

(57) 요약

본 발명은 다수 프레임의 스테레오 영상을 획득하는 스테레오 영상 획득부 및 동일 구조를 갖고 패턴 인식 방법이 미리 학습되어 동일한 학습 가중치를 갖는 다수의 컨볼루션 신경망(이하 CNN)을 포함하여, 스테레오 영상 획득부에서 전달되는 연속된 2개 프레임의 스테레오 영상 세트에서 연속된 프레임의 영상들에 대한 유틸리티플로우와 시차에 따라 구분되는 좌영상과 우영상 사이의 디스패리티를 동시에 추정하여 출력하는 추정부를 포함하고, 추정부의 다수의 CNN은 학습 시에 입력된 연속된 2개 프레임의 스테레오 영상 세트의 4개의 영상 중 하나의 영상의 각 픽셀로부터 기지정된 순방향 및 역방향 각각으로 나머지 영상을 사이클하여 탐색된 대응점의 위치의 변화의 합을 나타내는 사이클 전이 결과에 따라 획득되는 사이클 일관성 손실을 포함하는 총 손실이 역전파되어 업데이트된 학습 가중치로 학습된 유틸리티플로우 및 디스패리티 추정 장치 및 방법을 제공할 수 있다.

대표도 - 도9



(52) CPC특허분류

**H04N 13/128** (2018.05)

**H04N 13/204** (2018.05)

이 발명을 지원한 국가연구개발사업

과제고유번호 2018-11-1334

부처명 과학기술정보통신부

연구관리전문기관 한국연구재단

연구사업명 복합인지기술개발사업

연구과제명 [Ezbaro] (2세부)이중 CCTV 영상에서의 딥러닝 기반 실종자 초동 신원확인 및 추적 시스템  
(1단계)(1/2)

기 여 율 1/1

주관기관 연세대학교 산학협력단

연구기간 2018.10.23 ~ 2019.04.22

---

## 명세서

### 청구범위

#### 청구항 1

다수 프레임의 스테레오 영상을 획득하는 스테레오 영상 획득부; 및

동일 구조를 갖고 패턴 인식 방법이 미리 학습되어 동일한 학습 가중치를 갖는 다수의 컨볼루션 신경망(이하 CNN)을 포함하여, 상기 스테레오 영상 획득부에서 전달되는 연속된 2개 프레임의 스테레오 영상 세트에서 연속된 프레임의 영상들에 대한 옵티컬플로우와 시차에 따라 구분되는 좌영상과 우영상 사이의 디스패리티를 동시에 추정하여 출력하는 추정부; 를 포함하고,

상기 추정부의 다수의 CNN은

학습 시에 입력된 연속된 2개 프레임의 스테레오 영상 세트의 4개의 영상 중 하나의 영상의 각 픽셀로부터 기지정된 순방향 및 역방향 각각으로 나머지 영상을 사이클하여 탐색된 대응점의 위치의 변화의 합을 나타내는 사이클 전이 결과에 따라 획득되는 사이클 일관성 손실을 포함하는 총 손실이 역전파되어 업데이트된 상기 학습 가중치로 학습된 옵티컬플로우 및 디스패리티 추정 장치.

#### 청구항 2

제1 항에 있어서, 상기 사이클 일관성 손실은

상기 4개의 영상 중 순방향 및 역방향으로의 사이클 경로에서 2개씩의 영상 사이에 대응점의 존재 여부를 나타내는 사이클 신뢰도 맵을 상기 사이클 전이 결과에 반영하여 획득되는 옵티컬플로우 및 디스패리티 추정 장치.

#### 청구항 3

제2 항에 있어서, 상기 사이클 일관성 손실은

순방향 및 역방향으로의 각 픽셀에 대한 사이클 일관성 손실이 기지정된 문턱값을 초과하면, 해당 픽셀의 사이클 일관성 손실을 문턱값으로 출력하는 옵티컬플로우 및 디스패리티 추정 장치.

#### 청구항 4

제1 항에 있어서, 상기 총 손실은

학습 시에 상기 4개의 영상 중 프레임 순서에 따른 2개의 영상 사이, 동일 프레임의 2개의 영상 사이, 프레임과 시차가 상이한 2개의 영상 사이 각각에서 각 픽셀에 대한 대응점의 픽셀값과 그라디언트 값에 따라 획득되는 복원 손실을 추가로 포함하는 옵티컬플로우 및 디스패리티 추정 장치.

#### 청구항 5

제4 항에 있어서, 상기 복원 손실은

프레임 순서에 따른 2개의 영상 사이, 동일 프레임의 2개의 영상 사이, 프레임과 시차가 상이한 2개의 영상 사이 각각에서 대응점의 존재 여부를 나타내는 신뢰도 맵을 더 반영하여 획득되는 옵티컬플로우 및 디스패리티 추정 장치.

#### 청구항 6

제4 항에 있어서, 상기 총 손실은

상기 4개의 영상에서 프레임 순서에 따라 획득된 옵티컬플로우 변화를 제한하는 옵티컬플로우 평활화 손실과 시차에 따라 획득된 디스패리티의 변화를 제한하는 디스패리티 평활화 손실을 추가로 포함하는 옵티컬플로우 및 디스패리티 추정 장치.

#### 청구항 7

제6 항에 있어서, 상기 옵티컬플로우 및 디스패리티 추정 장치는

상기 추정부의 다수의 CNN을 학습시키는 동안 결합되어 상기 학습 가중치를 획득하는 학습부를 더 포함하고,

상기 학습부는

상기 추정부의 다수의 CNN과 동일한 구조를 갖고, 상기 4개의 영상 중 서로 다른 조합의 2개씩의 영상에 대해 대응점의 위치 변화를 오프셋으로 획득하는 다수의 삼 CNN으로 구성된 오프셋 획득부;

상기 오프셋 획득부의 다수의 삼 CNN 각각에서 획득되는 다수의 오프셋을 이용하여, 상기 사이클 일관성 손실, 상기 복원 손실, 상기 옵티컬플로우 평활화 손실 및 상기 디스패리티 평활화 손실을 계산하는 손실 측정부; 및

상기 사이클 일관성 손실, 상기 복원 손실, 상기 옵티컬플로우 평활화 손실 및 상기 디스패리티 평활화 손실 각각에 대해 기지정된 손실 가중치를 적용하여 상기 총 손실을 획득하여 상기 다수의 삼 CNN으로 역전파하여, 상기 다수의 삼 CNN에 대한 학습 가중치를 업데이트하고, 상기 다수의 삼 CNN에 대한 학습이 완료되면, 상기 학습 가중치를 상기 추정부의 다수의 CNN으로 전달하는 손실 역전파부; 를 포함하는 옵티컬플로우 및 디스패리티 추정 장치.

## 청구항 8

다수 프레임의 스테레오 영상을 획득하는 단계; 및

동일 구조를 갖고 패턴 인식 방법이 미리 학습되어 동일한 학습 가중치를 갖는 다수의 컨볼루션 신경망(이하 CNN)을 이용하여, 상기 다수 프레임의 스테레오 영상 중 연속된 2개 프레임의 스테레오 영상 세트에서 연속된 프레임의 영상들에 대한 옵티컬플로우와 시차에 따라 구분되는 좌영상과 우영상 사이의 디스패리티를 동시에 추정하여 출력하는 단계; 를 포함하고,

상기 다수의 CNN은

학습 시에 입력된 연속된 2개 프레임의 스테레오 영상 세트의 4개의 영상 중 하나의 영상의 각 픽셀로부터 기지정된 순방향 및 역방향 각각으로 나머지 영상을 사이클하여 탐색된 대응점의 위치의 변화의 합을 나타내는 사이클 전이 결과에 따라 획득되는 사이클 일관성 손실을 포함하는 총 손실이 역전파되어 업데이트된 상기 학습 가중치로 학습된 옵티컬플로우 및 디스패리티 추정 방법.

## 청구항 9

제8 항에 있어서, 상기 옵티컬플로우 및 디스패리티 추정 방법은

상기 다수의 CNN을 학습시키는 학습 단계를 더 포함하고,

상기 학습 단계는

상기 다수의 CNN과 동일한 구조를 갖는 다수의 삼 CNN을 이용하여, 학습 시에 입력된 상기 4개의 영상 중 서로 다른 조합의 2개씩의 영상에 대한 대응점의 위치 변화를 나타내는 다수의 오프셋으로 획득하는 단계;

상기 다수의 오프셋을 이용하여, 상기 사이클 전이 결과와 상기 4개의 영상 중 순방향 및 역방향으로의 사이클 경로에서 2개씩의 영상 사이에 대응점의 존재 여부를 나타내는 사이클 신뢰도 맵을 획득하고, 상기 사이클 신뢰도 맵을 상기 사이클 전이 결과에 반영하여 상기 사이클 일관성 손실을 획득하는 단계;

상기 4개의 영상 중 프레임 순서에 따른 2개의 영상 사이, 동일 프레임의 2개의 영상 사이, 프레임과 시차가 상이한 2개의 영상 사이 각각에서 각 픽셀에 대한 대응점의 픽셀값과 그래디언트 값에 따라 획득되는 복원 손실을 계산하는 단계;

상기 4개의 영상에서 프레임 순서에 따라 획득된 옵티컬플로우 변화를 제한하는 옵티컬플로우 평활화 손실과 시차에 따라 획득된 디스패리티의 변화를 제한하는 디스패리티 평활화 손실을 계산하는 단계;

상기 사이클 일관성 손실, 상기 복원 손실, 상기 옵티컬플로우 평활화 손실 및 상기 디스패리티 평활화 손실 각각에 대해 기지정된 손실 가중치를 적용하여 상기 총 손실을 획득하는 단계;

상기 총 손실을 상기 다수의 삼 CNN으로 역전파하여, 상기 다수의 삼 CNN에 대한 학습 가중치를 업데이트하는 단계; 및

상기 다수의 삼 CNN에 대한 학습이 완료되면, 상기 학습 가중치를 상기 다수의 CNN으로 전달하는 단계; 를 포함하는 옵티컬플로우 및 디스패리티 추정 방법.

## 청구항 10

제9 항에 있어서, 상기 학습 단계는

상기 4개의 영상 중 서로 다른 조합의 2개의 영상 사이 각각에서 대응점의 존재 여부를 나타내는 신뢰도 맵을 생성하는 단계; 및

상기 신뢰도 맵을 이용하여 상기 4개의 영상 중 순방향 및 역방향으로의 사이클 경로에서 2개씩의 영상 사이 대응점의 존재 여부를 나타내는 사이클 신뢰도 맵을 생성하는 단계; 를 더 포함하고,

상기 사이클 일관성 손실을 획득하는 단계는

상기 사이클 신뢰도 맵을 상기 사이클 전이 결과에 반영하여 상기 사이클 일관성 손실을 획득하고,

상기 복원 손실을 계산하는 단계는

프레임 순서에 따른 2개의 영상 사이, 동일 프레임의 2개의 영상 사이, 프레임과 시차가 상이한 2개의 영상 사이 각각에서 획득된 신뢰도 맵을 반영하여 상기 복원 손실을 획득하는 옵티컬플로우 및 디스패리티 추정 방법.

## 발명의 설명

### 기술 분야

[0001] 본 발명은 옵티컬플로우 및 디스패리티 추정 장치 및 방법에 관한 것으로, 사이클 일관성에 기반한 옵티컬플로우 및 디스패리티 동시 추정 장치 및 방법에 관한 것이다.

### 배경 기술

[0002] 다수의 이미지 사이의 밀집 대응점(dense correspondence) 분석은 영상 처리(image processing) 및 컴퓨터 비전(computer vision) 분야의 다양한 어플리케이션에 있어서 기본적인 작업으로, 특히 첨단 운전자 보조 시스템(Advanced Driver Assistance System: ADAS) 및 자율 주행 시스템(autonomous driving system)을 포함한 다양한 차량 어플리케이션에서 주로 이용되고 있다.

[0003] 이미지의 대응점 분석은 이미지 스티칭(image stitching), 정렬(alignment), 인식(recognition), 스테레오 매칭(stereo matching), 및 옵티컬플로우(optical flow) 등에서 주로 이용되고 있다.

[0004] 한편, 인공 신경망을 이용한 딥러닝 기법에 대한 연구가 발전됨에 따라 스테레오 매칭을 위한 디스패리티 추정 및 옵티컬플로우 추정에도 딥러닝 기법이 이용되는 비중이 높아지고 있다. 딥러닝 기법을 이용한 디스패리티 추정 및 옵티컬플로우 추정에 대한 많은 연구가 수행되었지만, 기본적으로 스테레오 영상의 매칭을 위한 디스패리티 추정과 연속되는 다수 프레임에서 객체의 움직임을 유추하기 위한 옵티컬플로우 추정은 독자적으로 연구가 수행되어 왔다.

[0005] 그러나 최근에 딥러닝 기법을 이용하여 디스패리티와 옵티컬플로우를 동시에 학습하여 추정하는 경우, 디스패리티와 옵티컬플로우를 개별적으로 추정하는 것보다 성능이 향상되는 것으로 밝혀졌다. 그러나 기존에는 디스패리티와 옵티컬플로우를 개별로 추정하기 위해 학습하는 경우나 디스패리티와 옵티컬플로우를 동시에 추정하기 위해 학습하는 경우 모두 기본적으로 영상에 검증 자료(ground truth)가 레이블로 포함된 학습 데이터를 이용한 지도 학습(supervised learning) 방식을 기반으로 연구되어 왔다.

[0006] 현재 디스패리티와 옵티컬플로우를 개별로 지도 학습시키기 위한 학습 데이터도 부족하지만, 디스패리티와 옵티컬플로우를 동시에 지도 학습시키기 위한 학습 데이터는 특히 부족한 실정이다. 그리고 학습 데이터의 부족은 옵티컬플로우 및 디스패리티 추정 시에 정확도를 크게 떨어뜨리는 결과를 초래하는 문제가 있다.

[0007] 그러나 레이블된 검증 자료가 포함된 학습 데이터의 생성은 매우 오랜 시간과 노력 및 비용이 요구된다. 이에 가상의 이미지와 검증 자료를 합성한 가상 학습 데이터를 생성하는 방식이 연구되었으나, 가상 학습 데이터를 이용하여 학습된 경우, 가상 영상의 사실성과 가변성의 결여로 인해 요구되는 성능을 나타내지 못하는 경우가 대부분이다.

[0008] 따라서 옵티컬플로우 및 디스패리티 추정 장치가 검증 자료가 레이블된 학습 데이터를 요구하지 않고, 통상의 스테레오 영상으로부터 비지도 학습 방식으로 디스패리티와 옵티컬플로우를 동시에 추론 가능하도록 학습되는 기법이 요구되고 있다.

## 선행기술문헌

### 특허문헌

[0009] (특허문헌 0001) 한국 등록 특허 제10-1849605호 (2018.04.11 등록)

## 발명의 내용

### 해결하려는 과제

[0010] 본 발명의 목적은 학습 데이터를 요구하지 않는 비지도 학습 방식으로 학습되어도 높은 옵티컬플로우 및 디스패리티의 동시 추정 성능을 나타내는 옵티컬플로우 및 디스패리티 추정 장치 및 방법을 제공하는데 있다.

[0011] 본 발명의 다른 목적은 옵티컬플로우 및 디스패리티 추정을 위한 학습을 동시에 수행할 수 있어, 학습 시간을 저감할 수 있는 옵티컬플로우 및 디스패리티 추정 장치 및 방법을 제공하는데 있다.

[0012] 본 발명의 또 다른 목적은 검증 자료가 포함되지 않은 연속하는 다수 프레임의 스테레오 영상에서 연속 프레임의 좌영상 및 우영상 간 대응점의 사이클 일관성에 기반하여 비지도 학습되는 옵티컬플로우 및 디스패리티 추정 장치 및 방법을 제공하는데 있다.

### 과제의 해결 수단

[0013] 상기 목적을 달성하기 위한 본 발명의 일 실시예에 따른 옵티컬플로우 및 디스패리티 추정 장치는 다수 프레임의 스테레오 영상을 획득하는 스테레오 영상 획득부; 및 동일 구조를 갖고 패턴 인식 방법이 미리 학습되어 동일한 학습 가중치를 갖는 다수의 컨볼루션 신경망(이하 CNN)을 포함하여, 상기 스테레오 영상 획득부에서 전달되는 연속된 2개 프레임의 스테레오 영상 세트에서 연속된 프레임의 영상들에 대한 옵티컬플로우와 시차에 따라 구분되는 좌영상과 우영상 사이의 디스패리티를 동시에 추정하여 출력하는 추정부; 를 포함하고, 상기 추정부의 다수의 CNN은 학습 시에 입력된 연속된 2개 프레임의 스테레오 영상 세트의 4개의 영상 중 하나의 영상의 각 픽셀로부터 기지정된 순방향 및 역방향 각각으로 나머지 영상을 사이클하여 탐색된 대응점의 위치의 변화의 합을 나타내는 사이클 전이 결과에 따라 획득되는 사이클 일관성 손실을 포함하는 총 손실이 역전파되어 업데이트된 상기 학습 가중치로 학습될 수 있다.

[0014] 상기 사이클 일관성 손실은 상기 4개의 영상 중 순방향 및 역방향으로의 사이클 경로에서 2개씩의 영상 사이에 대응점의 존재 여부를 나타내는 사이클 신뢰도 맵을 상기 사이클 전이 결과에 반영하여 획득될 수 있다.

[0015] 상기 사이클 일관성 손실은 순방향 및 역방향으로의 각 픽셀에 대한 사이클 일관성 손실이 기지정된 문턱값을 초과하면, 해당 픽셀의 사이클 일관성 손실을 문턱값으로 출력할 수 있다.

[0016] 상기 총 손실은 학습 시에 상기 4개의 영상 중 프레임 순서에 따른 2개의 영상 사이, 동일 프레임의 2개의 영상 사이, 프레임과 시차가 상이한 2개의 영상 사이 각각에서 각 픽셀에 대한 대응점의 픽셀값과 그래디언트 값에 따라 획득되는 복원 손실을 추가로 포함할 수 있다.

[0017] 상기 복원 손실은 프레임 순서에 따른 2개의 영상 사이, 동일 프레임의 2개의 영상 사이, 프레임과 시차가 상이한 2개의 영상 사이 각각에서 대응점의 존재 여부를 나타내는 신뢰도 맵을 더 반영하여 획득할 수 있다.

[0018] 상기 총 손실은 상기 4개의 영상에서 프레임 순서에 따라 획득된 옵티컬플로우 변화를 제한하는 옵티컬플로우 평활화 손실과 시차에 따라 획득된 디스패리티의 변화를 제한하는 디스패리티 평활화 손실을 추가로 포함할 수 있다.

[0019] 상기 옵티컬플로우 및 디스패리티 추정 장치는 상기 추정부의 다수의 CNN을 학습시키는 동안 결합되어 상기 학습 가중치를 획득하는 학습부를 더 포함하고, 상기 학습부는 상기 추정부의 다수의 CNN과 동일한 구조를 갖고, 상기 4개의 영상 중 서로 다른 조합의 2개씩의 영상에 대해 대응점의 위치 변화를 오프셋으로 획득하는 다수의 삼 CNN으로 구성된 오프셋 획득부; 상기 오프셋 획득부의 다수의 삼 CNN 각각에서 획득되는 다수의 오프셋을 이

용하여, 상기 사이클 일관성 손실, 상기 복원 손실, 상기 옵티컬플로우 평활화 손실 및 상기 디스패리티 평활화 손실을 계산하는 손실 측정부; 및 상기 사이클 일관성 손실, 상기 복원 손실, 상기 옵티컬플로우 평활화 손실 및 상기 디스패리티 평활화 손실 각각에 대해 기지정된 손실 가중치를 적용하여 상기 총 손실을 획득하여 상기 다수의 삼 CNN으로 역전파하여, 상기 다수의 삼 CNN에 대한 학습 가중치를 업데이트하고, 상기 다수의 삼 CNN에 대한 학습이 완료되면, 상기 학습 가중치를 상기 추정부의 다수의 CNN으로 전달하는 손실 역전파부; 를 포함할 수 있다.

[0020] 상기 목적을 달성하기 위한 본 발명의 다른 실시예에 따른 옵티컬플로우 및 디스패리티 추정 방법은 다수 프레임의 스테레오 영상을 획득하는 단계; 및 동일 구조를 갖고 패턴 인식 방법이 미리 학습되어 동일한 학습 가중치를 갖는 다수의 컨볼루션 신경망(이하 CNN)을 이용하여, 상기 다수 프레임의 스테레오 영상 중 연속된 2개 프레임의 스테레오 영상 세트에서 연속된 프레임의 영상들에 대한 옵티컬플로우와 시차에 따라 구분되는 좌영상과 우영상 사이의 디스패리티를 동시에 추정하여 출력하는 단계; 를 포함하고, 상기 다수의 CNN은 학습 시에 입력된 연속된 2개 프레임의 스테레오 영상 세트의 4개의 영상 중 하나의 영상의 각 픽셀로부터 기지정된 순방향 및 역방향 각각으로 나머지 영상을 사이클하여 탐색된 대응점의 위치의 변화의 합을 나타내는 사이클 전이 결과에 따라 획득되는 사이클 일관성 손실을 포함하는 총 손실이 역전파되어 업데이트된 상기 학습 가중치로 학습될 수 있다.

### 발명의 효과

[0021] 따라서, 본 발명의 실시예에 따른 옵티컬플로우 및 디스패리티 추정 장치 및 방법은 연속하는 프레임의 스테레오 영상에서 대응점에 대한 사이클이 동일해야 하는 사이클 일관성을 이용하여, 옵티컬플로우와 디스패리티를 동시에 비지도 학습 방식으로 학습시킬 수 있으므로, 검증 자료가 포함된 학습 데이터를 요구하지 않고 비지도 학습 방식으로 학습되어 매우 높은 옵티컬플로우와 디스패리티 추정 성능을 가지며, 학습 시간을 크게 줄일 수 있다. 또한 신뢰도 맵에 기반하여 사이클 일관성에 기초한 학습시의 정확도를 높일 수 있으며, 복원 손실 및 평활화 손실이 추가로 역전파되어 학습됨으로써, 옵티컬플로우와 디스패리티 추정 성능을 더욱 향상시킬 수 있다.

### 도면의 간단한 설명

[0022] 도1 은 본 발명의 일 실시예에 따른 옵티컬플로우 및 디스패리티 추정 장치의 개략적 구조를 나타낸다.  
 도2 는 도1 의 학습부의 상세 구성을 나타낸다.  
 도3 은 도2 의 오프셋 획득부의 동작을 설명하기 위한 도면이다.  
 도4 는 도2 의 손실 측정부의 상세 구성을 나타낸다.  
 도5 는 사이클 일관성 손실의 개념을 설명하기 위한 도면이다.  
 도6 은 신뢰도 맵의 개념을 설명하기 위한 도면이다.  
 도7 및 도8 은 복원 손실의 개념을 설명하기 위한 도면이다.  
 도9 는 본 발명의 일 실시예에 따른 옵티컬플로우 및 디스패리티 추정 방법을 나타낸다.  
 도10 및 도11 은 본 실시예에 따른 옵티컬플로우 및 디스패리티 추정 장치 및 방법의 성능을 비교한 결과를 나타낸다.

### 발명을 실시하기 위한 구체적인 내용

[0023] 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.

[0024] 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.

[0025] 명세서 전체에서, 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라, 다른 구성요소를 더 포함할 수 있는 것을 의미한다. 또한, 명세서에 기



재된 "...부", "...기", "모듈", "블록" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.

- [0026] 도1 은 본 발명의 일 실시예에 따른 옵티컬플로우 및 디스패리티 추정 장치의 개략적 구조를 나타낸다.
- [0027] 도1 을 참조하면, 본 실시예의 옵티컬플로우 및 디스패리티 추정 장치는 스테레오 영상 획득부(100), 추정부(200) 및 학습부(300)를 포함한다.
- [0028] 스테레오 영상 획득부(100)는 옵티컬플로우 및 디스패리티를 추정해야할 스테레오 영상을 획득한다. 여기서 스테레오 영상은 스테레오 카메라에서 획득될 수 있는 영상으로 2개의 서로 다른 시점을 갖는 다수의 연속된 프레임으로 구성된 영상이다. 스테레오 영상은 스테레오 카메라의 구조에 따라 상하 영상 좌우 영상 등으로 획득될 수 있으나, 여기서는 일예로 좌영상(Left image)과 우영상(Right image)을 획득하는 것으로 가정한다.
- [0029] 스테레오 영상 획득부(100)는 획득되는 다수 프레임의 스테레오 영상 중 연속된 2개 프레임의 스테레오 영상을 추정부(200)로 전달한다.
- [0030] 스테레오 영상 획득부(100)는 일예로 t번째 프레임 및 t+1번째 프레임을 추정부(200)로 전달한다. 즉 t번째 프레임의 좌영상 및 우영상, 즉 스테레오 영상 세트( $l_1, r_1$ )와 t+1번째 프레임의 스테레오 영상 세트( $l_2, r_2$ )를 전달한다.
- [0031] 추정부(200)는 각각 미리 학습된 인공 신경망, 일예로 컨볼루션 신경망(Convolution neural network: 이하 CNN)으로 구현되는 옵티컬플로우 추정부(210) 및 디스패리티 추정부(220)를 포함하여, 스테레오 영상 획득부(100)에서 전달된 2개의 스테레오 영상 세트로부터 옵티컬플로우 및 디스패리티를 추정하여 출력한다.
- [0032] 옵티컬플로우 추정부(210)는 연속된 2개의 프레임의 좌영상 및 우영상 각각에서 대응점의 이동, 즉 객체의 움직임 탐색하여 옵티컬플로우를 추정하기 위한 2개의 CNN을 포함할 수 있으며, 디스패리티 추정부(220)는 각 프레임의 좌영상과 우영상 사이의 대응점 위치 차인 디스패리티를 추정하기 위한 적어도 하나의 CNN을 포함할 수 있다.
- [0033] 여기서 옵티컬플로우 추정부(210)와 디스패리티 추정부(220)에 포함된 다수의 CNN은 동일한 구조를 갖고, 동시에 학습되어 동일한 가중치가 적용되는 삼(siamese) CNN으로 구현될 수 있다.
- [0034] 비록 옵티컬플로우는 연속하는 프레임에서 객체의 움직임을 탐색하는 반면, 디스패리티는 동일 프레임의 좌영상 및 우영상의 차이를 탐색하는 차이가 있으나, 기본적으로 2개의 영상에서 대응하는 픽셀 간의 위치 변이, 즉 대응점 변이를 추정한다는 점에서 유사하다. 따라서, 동일 구조를 갖고 동시에 학습되어 동일한 가중치가 적용된 삼 CNN으로 구현될 수 있다.
- [0035] 옵티컬플로우 추정부(210)는 2개의 CNN을 포함하여, 스테레오 영상 획득부(100)에서 전달된 2개의 스테레오 영상 세트( $(l_1, r_1), (l_2, r_2)$ )의 t번째 프레임 및 t+1번째 프레임의 좌영상들( $l_1, l_2$ ) 사이의 대응점 변이( $F_{l_1, l_2}$ )와 우영상들( $r_1, r_2$ ) 사이의 대응점 변이( $F_{r_1, r_2}$ )를 옵티컬플로우로 추정한다.
- [0036] 그리고 디스패리티 추정부(220)는 스테레오 영상 획득부(100)에서 전달된 2개의 스테레오 영상 세트( $(l_1, r_1), (l_2, r_2)$ )의 t번째 프레임의 좌영상( $l_1$ )에서 우영상( $r_1$ )으로의 대응점 변이( $F_{l_1, r_1}$ )를 디스패리티로 추정한다. 이때 디스패리티 추정부(220)는 다수의 CNN을 포함하여 t+1번째 프레임의 좌영상( $l_2$ )에서 우영상( $r_2$ )으로의 대응점 변이( $F_{l_2, r_2}$ ) 또한 함께 추정할 수 있다. 또한 경우에 따라서는 t번째 프레임 및 t+1번째 프레임 각각의 우영상( $r_1, r_2$ )에서 좌영상( $l_1, l_2$ )으로의 대응점 변이( $(F_{r_1, l_1}), (F_{r_2, l_2})$ )도 추정할 수도 있다.
- [0037] 상기한 바와 같이, 본 실시예에 따른 옵티컬플로우 및 디스패리티 추정 장치에서 추정부(200)는 미리 학습된 다수의 삼 CNN으로 구성되어, 전달된 2개 프레임의 스테레오 영상으로부터 옵티컬플로우 및 디스패리티를 동시에 추정하여 출력할 수 있다.
- [0038] 다만 추정부(200)의 다수의 삼 CNN이 옵티컬플로우 및 디스패리티를 정확하게 추정하기 위해서는 반드시 미리 학습된 상태이어야 한다.
- [0039] 이에 실시예에 따른 옵티컬플로우 및 디스패리티 추정 장치는 별도의 검증 자료가 레이블된 학습 데이터 없이 추정부(200)를 학습시키기 위한 학습부(300)를 더 포함할 수 있다.



- [0040] 학습부(300)는 추정부(200)와 마찬가지로 스테레오 영상 획득부(100)로부터 2개 프레임의 스테레오 영상 세트  $((l_1, r_1), (l_2, r_2))$ 를 전달받는다. 그리고 전달된 2개 프레임의 스테레오 영상 세트  $((l_1, r_1), (l_2, r_2))$ 에 대해 추정부(200)와 동일한 구성을 갖는 8개 삼 CNN을 이용하여 8개의 오프셋을 획득하고, 획득된 8개의 오프셋으로부터 손실을 측정하여 8개의 삼 CNN으로 역전파함으로써 8개의 삼 CNN의 가중치를 업데이트 한다. 즉 8개의 삼 CNN을 학습시킨다.
- [0041] 여기서 8개의 삼 CNN은 추정부(200)와 동일한 구성을 가지므로, 업데이트된 8개의 삼 CNN의 가중치는 추정부(200)에 포함된 다수의 삼 CNN의 가중치로 적용되며, 결과적으로 추정부(200)의 다수의 삼 CNN을 학습시킬 수 있다.
- [0042] 그리고 학습부(300)는 학습의 성능을 향상시키기 위해, 즉 추정부(200)의 옵티컬플로우 및 디스패리티 추정 정확도를 높이기 위해, 8개의 오프셋으로부터 여러 종류의 손실을 측정하여 역전파할 수 있다. 여기서는 일예로 사이클 일관성 손실, 복원 손실 및 평활화 손실 측정하여 역전파한다.
- [0043] 학습부(300)는 단지 추정부(200)를 학습시키기 위한 구성으로, 추정부의 다수의 삼 CNN이 학습된 이후에는 생략될 수 있다. 즉 옵티컬플로우 및 디스패리티 추정 장치가 실제 이용되는 경우에는 학습부(300)가 제외될 수 있다. 뿐만 아니라, 학습부(300)는 다수의 삼 CNN을 위한 가중치를 획득하는 별도의 학습 장치로 구성될 수도 있다.
- [0044] 도2 는 도1 의 학습부의 상세 구성을 나타내고, 도3 은 도2 의 오프셋 획득부의 동작을 설명하기 위한 도면이며, 도4 는 도2 의 손실 측정부의 상세 구성을 나타낸다.
- [0045] 도2 를 참조하면, 학습부(300)는 오프셋 획득부(310), 손실 측정부(320) 및 손실 역전파부(330)를 포함한다.
- [0046] 오프셋 획득부(310)는 도3 에 도시된 바와 같이, 각각 스테레오 영상 획득부(100)에서 전달된 2개의 스테레오 영상 세트  $((l_1, r_1), (l_2, r_2))$ 에서 각각 대응하는 2개의 영상을 인가받아 특징 맵을 추출하는 인코더(Encoder)와 추출된 특징 맵으로부터 오프셋을 획득하는 디코더(Decoder) 구조를 가지는 8개의 삼 CNN을 포함한다. 그리고 8개의 삼 CNN 중 4개의 삼 CNN은 옵티컬플로우 오프셋 획득부(311)를 구성하고, 나머지 4개의 삼 CNN은 디스패리티 오프셋 획득부(312)를 구성한다.
- [0047] 표1 은 삼 CNN의 인코더 및 디코더 구조의 일예를 나타낸다.

표 1

Encoder	Kernel	Stride	Ch I/O	Input	Decoder	Kernel	Stride	Ch I/O	Input
conv1	7×7	2	6/64	images	upconv6	4×4	2	1024/512	conv6b
conv2	5×5	2	64/128	conv1	conv6d	3×3	1	1024/1024	conv5b+upconv6
conv3a	5×5	2	128/256	conv2	upconv5	4×4	2	1024/256	conv6d
conv3b	3×3	1	256/256	conv3a	conv5d	3×3	1	768/768	conv4b+upconv5
conv4a	3×3	2	256/512	conv3b	upconv4	4×4	2	768/128	conv5d
conv4b	3×3	1	512/512	conv4a	conv4d	3×3	1	384/384	conv3b+upconv4
conv5a	3×3	2	512/512	conv4b	upconv3	4×4	2	384/64	conv4d
conv5b	3×3	1	512/512	conv5a	conv3d	3×3	1	192/192	conv2+upconv3
conv6a	3×3	2	512/1024	conv5b	upconv2	4×4	2	192/32	conv3d
conv6b	3×3	1	1024/1024	conv6a	conv2d	3×3	1	96/96	conv1+upconv2
					upconv1	4×4	2	96/16	conv2d
					conv1d	3×3	1	16/2	upconv1

- [0048]
- [0049] 표1 에 나타난 바와 같이, 인코더는 일예로 10개의 컨볼루션 레이어를 포함하여 구성될 수 있으며, 디코더는 컨볼루션 레이어와 업컨볼루션 레이어가 조합된 12개의 레이어를 포함할 수 있다.
- [0050] 옵티컬플로우 오프셋 획득부(311)는 동일한 구조의 4개의 삼 CNN을 포함하며, 각각의 CNN은 스테레오 영상 획득부(100)에서 전달된 2개의 스테레오 영상 세트  $((l_1, r_1), (l_2, r_2))$ 의 좌영상  $(l_1, l_2)$ 과 우영상  $(r_1, r_2)$  각각에 대해 시간의 흐름에 따라 순방향 및 역방향의 4개의 옵티컬플로우 오프셋  $((F_{l1,l2}), (F_{r2,r1}), (F_{l2,l1}), (F_{r1,r2}))$ 을 획득한다.
- [0051] 옵티컬플로우 오프셋 획득부(311)의 4개의 삼 CNN 중 2개는 추정부(200)의 옵티컬플로우 추정부(210)와 동일하게 스테레오 영상 획득부(100)에서 전달된 2개의 스테레오 영상 세트  $((l_1, r_1), (l_2, r_2))$ 에서 시간의 흐름에 따

라 순방향으로  $t$ 번째 프레임으로부터  $t+1$ 번째 프레임으로의 좌영상들( $l_1, l_2$ ) 사이의 대응점 변이( $F_{l_1, l_2}$ )와 우영상들( $r_1, r_2$ ) 사이의 대응점 변이( $F_{r_1, r_2}$ )를 옵티컬플로우 오프셋으로 획득한다.

[0052] 그리고 옵티컬플로우 오프셋 획득부(311)의 나머지 2개의 삼 CNN은 시간의 역방향으로  $t+1$ 번째 프레임으로부터  $t$ 번째 프레임으로의 좌영상들( $l_1, l_2$ ) 사이의 대응점 변이( $F_{l_2, l_1}$ )와 우영상들( $r_2, r_1$ ) 사이의 대응점 변이( $F_{r_1, r_2}$ )를 옵티컬플로우 오프셋으로 획득한다.

[0053] 한편, 디스패리티 오프셋 획득부(312) 또한 4개의 삼 CNN을 포함하고, 각각의 CNN은 스테레오 영상 획득부(100)에서 전달된 2개의 스테레오 영상 세트( $(l_1, r_1), (l_2, r_2)$ )에서  $t$ 번째 및  $t+1$ 번째 프레임 각각의 좌영상( $l_1, l_2$ )에서 우영상( $r_1, r_2$ )으로의 대응점 변이( $(F_{l_1, r_1}), (F_{l_2, r_2})$ ) 및 우영상( $r_1, r_2$ )에서 좌영상( $l_1, l_2$ )으로의 대응점 변이( $(F_{r_1, l_1}), (F_{r_2, l_2})$ )를 디스패리티 오프셋으로 획득한다.

[0054] 즉 오프셋 획득부(310)는 8개의 삼 CNN을 이용하여 스테레오 영상 획득부(100)에서 전달된 2개의 스테레오 영상 세트( $(l_1, r_1), (l_2, r_2)$ )의 4개의 영상에 대해 서로 다른 조합으로 대응점 변이를 탐색하여, 8개의 오프셋(4개의 옵티컬플로우 오프셋( $(F_{l_1, l_2}), (F_{r_2, r_1}), (F_{l_2, l_1}), (F_{r_1, r_2})$ )과 4개의 디스패리티 오프셋( $(F_{l_1, r_1}), (F_{r_2, l_2}), (F_{r_1, l_1}), (F_{l_2, r_2})$ ))을 획득한다.

[0055] 여기서  $a$  영상에서 픽셀( $p$ )의 위치가  $(p_x, p_y)$ 이고,  $b$  영상의 대응점( $q$ )의 위치가  $(q_x, q_y)$ 인 경우,  $a$  영상으로부터  $b$  영상으로의 대응점 변이, 즉 오프셋( $F_{a, b}$ )은 수학식 1과 같이 계산된다.

### 수학식 1

$$F_{a, b}(p) = (q_x - p_x, q_y - p_y)$$

[0056]

[0057] 본 실시예에서 오프셋 획득부(310)가 8개의 삼 CNN을 포함하여 8개의 오프셋을 획득하는 것은 학습부(300)가 별도의 학습 데이터 없이도 후술하는 사이클 일관성에 기반하여 비지도 학습을 정확하게 수행할 수 있도록 하기 위함이다.

[0058] 도4 를 참조하면, 손실 측정부(320)는 사이클 일관성 손실 측정부(321), 복원 손실 측정부(322), 평활화 손실 측정부(323) 및 신뢰도 맵 생성부(324)를 포함할 수 있다.

[0059] 상기한 바와 같이, 본 실시예에서 손실 측정부(320)는 오프셋 획득부(310)의 8개의 삼 CNN을 학습시키기 위해 사이클 일관성 손실, 복원 손실 및 평활화 손실의 3가지 손실을 측정한다.

[0060] 우선 사이클 일관성 손실 측정부(321)는 스테레오 영상 획득부(100)에서 전달된 2개 프레임의 스테레오 영상 세트( $(l_1, r_1), (l_2, r_2)$ )의 4개의 영상에서 기지점된 순서로 사이클된 대응점의 변이의 합이 0의 오프셋을 가져야 한다는 사이클 일관성을 이용하여 손실을 측정한다. 그리고 복원 손실 측정부(322)는 각 대응점의 픽셀 값은 유사해야 한다는 픽셀 일치성을 이용하여 손실을 측정한다. 마지막으로 평활화 손실 측정부(323)는 움직임 및 디스패리티의 경계를 제외한 나머지 영역에서 대응점의 주변 픽셀들의 픽셀 값은 매끄럽게 변화되어야 한다는 평활 특성을 이용하여 손실을 측정한다.

[0061] 한편, 신뢰도 맵 생성부(324)는 사이클 일관성 손실 측정부(321)와 복원 손실 측정부(322)가 객체의 움직임으로 인해 사라진 영역 및 폐색(occlusion) 영역과 같이 대응점이 존재하지 않는 영역에 대해 손실을 측정함으로써, 손실 측정의 오차가 발생하는 것을 방지하기 위해 신뢰도 맵을 생성한다.

[0062] 이하에서는 도5 내지 도9 를 참조하여, 도4 의 손실 측정부(320)의 동작을 상세하게 설명한다.

[0063] 도5 는 사이클 일관성 손실의 개념을 설명하기 위한 도면이다.

[0064] 사이클 일관성은 연속되는 2개의 프레임의 스테레오 영상 세트의 4개의 영상( $(l_1, r_1), (l_2, r_2)$ )에서 사이클되는 대응점은 일치되어야 함을 의미한다. 만일 오프셋 획득부(310)의 8개의 삼 CNN이 정상적으로 학습된 상태라고 가정하면, 도5 의 (a)에 도시된 바와 같이,  $t$ 번째 프레임의 좌영상( $l_1$ )의 특정 픽셀로부터  $t+1$ 번째 프레임의

좌영상( $l_2$ ),  $t+1$ 번째 프레임의 우영상( $r_2$ ) 및  $t$ 번째 프레임의 우영상( $r_1$ )을 거쳐 사이클된  $t$ 번째 프레임의 좌영상( $l_1$ )의 대응점은 최초  $t$ 번째 프레임의 좌영상( $l_1$ )의 픽셀과 동일해야 한다. 즉 오프셋 획득부(310)에서 획득된 각 대응점 변이의 합은 0이되어야 한다.

[0065] 그리고 이러한 사이클 일관성은 도5 의(a)에 도시된 순방향( $l_1 \rightarrow l_2 \rightarrow r_2 \rightarrow r_1$ )뿐만 아니라, (b)에 도시된 역방향( $l_1 \rightarrow r_1 \rightarrow r_2 \rightarrow l_2$ )에서도 만족해야만 한다.

[0066] 만일 순방향 및 역방향 사이클 중 적어도 하나에서 대응점 변이의 합이 0이 아니면, 이는 8개의 CNN 중 적어도 하나의 CNN이 잘못된 대응점에 대한 변이를 획득한 것으로 판단할 수 있으며, 사이클 일관성 손실로 나타나게 된다.

[0067] 이에 사이클 일관성 손실 측정부(321)는 순방향 및 역방향의 대응점 변이의 합을 계산한다. 그리고 대응점 변이의 합을 계산하기 위해 본 실시예에서는 우선 전이 연산자( $\circ$ )를 수학식 2과 같이 정의한다.

### 수학식 2

[0068] 
$$(F_{a,b} \circ F_{b,c})(p) = F_{a,b}(p) + F_{b,c}(p + F_{a,b}(p))$$

[0069] 수학식 2는 전이 연산자( $\circ$ )는  $a$ ,  $b$  및  $c$ 의 3개의 영상에 대해  $a$  영상으로부터  $b$  영상을 거쳐  $c$  영상으로의 전이되는 오프셋( $F_{a,b}$ )의 합을 나타낸다.

[0070] 그리고 도5 와 유사하게  $a$ ,  $b$ ,  $c$ ,  $d$ 의 4개의 영상에 대해 순차적으로 전이되는 사이클 전이 결과( $\bar{F}_{a,b}$ )를 수학식 2의 전이 연산자( $\circ$ )를 이용하면, 수학식 3과 같이 표현될 수 있다.

### 수학식 3

[0071] 
$$\bar{F}_{a,b} = F_{a,b} \circ F_{b,c} \circ F_{c,d} \circ F_{d,a}$$

[0072] 수학식 3을 참조하면,  $\bar{F}_{a,b}$ 는  $a$ ,  $b$ ,  $c$ ,  $d$ 의 4개의 영상에 대해  $a$  영상으로부터  $b$  영상 방향으로의 사이클 전이 결과를 나타낸다.

[0073] 이에 수학식 3을 도5 의 (a) 및 (b)에 반영하면, (a)의 순방향 사이클 전이 결과는  $\bar{F}_{l_1,l_2}$ 로 표현되고, (b)의 역방향 사이클 전이 결과는  $\bar{F}_{l_1,r_1}$ 로 표현될 수 있다.

[0074] 사이클 일관성을 만족하기 위해서는 순방향 사이클 전이 결과( $\bar{F}_{l_1,l_2}$ )와 역방향 사이클 전이 결과( $\bar{F}_{l_1,r_1}$ )가 모두 0이 되어야 하며, 전이 결과가 0이 아니면, 사이클 일관성 손실이 발생된 것으로 볼 수 있다. 즉 사이클 일관성 손실( $L_c$ )은 순방향 사이클 전이 결과( $\bar{F}_{l_1,l_2}$ )와 역방향 사이클 전이 결과( $\bar{F}_{l_1,r_1}$ )의 합( $\bar{F}_{l_1,l_2} + \bar{F}_{l_1,r_1}$ )으로 계산될 수 있다.

[0075] 그러나 상기한 바와 같이 2개의 스테레오 영상 세트의 4개의 영상( $(l_1, r_1)$ ,  $(l_2, r_2)$ ) 사이에는 대응점이 존재하지 않는 픽셀이 포함될 수 있다. 따라서 영상의 모든 픽셀( $p$ )에 대한 순방향 사이클 전이 결과( $\bar{F}_{l_1,l_2}$ )와 역방향 사이클 전이 결과( $\bar{F}_{l_1,r_1}$ )의 합을 단순 계산하는 경우, 계산된 사이클 일관성 손실( $L_c$ )에 오류가 발생하게 된다.

[0076] 이에 손실 측정부(320)는 신뢰도 맵 생성부(324)를 추가로 구비하여, 사이클 일관성 손실( $L_c$ )의 오류를 방지할

수 있도록 한다. 신뢰도 맵 생성부(324)는 2개의 스테레오 영상 세트의 4개의 영상( $(l_1, r_1)$ ,  $(l_2, r_2)$ )에서 각 조합에 따른 두 영상들 사이에서 획득되는 양방향 반복 대응점 변이 차, 즉 오프셋의 차를 기반으로 각 픽셀의 신뢰도를 판별한다.

[0077] 도6 은 신뢰도 맵의 개념을 설명하기 위한 도면으로, 도6 에서 (a)와 (b)는 a 영상(여기서는  $l_1$ )과 b 영상(여기서는  $l_2$ )의 두영상 사이에서 생성되는 신뢰도 맵의 개념을 나타낸다.

[0078] 신뢰도 맵 생성부(324)는 (a)에 도시된 a 영상에서 b 영상으로 전이 후 다시 a 영상으로 전이된 오프셋의 합( $F_{a,b} \circ F_{b,a}$ )과, (b)에 도시된 b 영상에서 a 영상으로 전이 후 다시 b 영상으로 전이된 오프셋의 합( $F_{b,a} \circ F_{a,b}$ ) 사이의 차를 기반으로 a 영상과 b 영상 사이의 신뢰도 맵을 계산한다.

[0079] 도6 을 참조하면, 오프셋 획득부(310)의 CNN들이 정상적으로 학습된 상태인 경우, (a)에 도시된 오프셋의 합( $F_{a,b} \circ F_{b,a}$ )과 (b)에 도시된 오프셋의 합( $F_{b,a} \circ F_{a,b}$ )은 모두 0으로 나타나야 하며, 적어도 0에 근접한 값을 나타내야 한다.

[0080] 따라서 오프셋의 합( $F_{a,b} \circ F_{b,a}$ )과 오프셋의 합( $F_{b,a} \circ F_{a,b}$ ) 사이의 차 또한 0에 근사된 값을 가져야 한다. 이에 a 영상과 b 영상의 두 영상 사이의 신뢰도 맵( $M_{a,b}$ )은 수학적 식 4에 따라 계산될 수 있다.

#### 수학적 식 4

$$M_{a,b} = \frac{H(\|F_{b,a} \circ F_{a,b}\|_2 - \|F_{a,b} \circ F_{b,a}\|_2 + \alpha_1)}{H(\alpha_2 - \|F_{a,b} \circ F_{b,a}\|_2)}$$

[0081]

[0082] 수학적 식 4 에서  $H(x)$ 는  $x$ 가 0 이상이면 1을 출력하고 이외에는 0을 출력하는 스텝 함수이고,  $\| \cdot \|_2$ 는 L2 norm 함수이다. 그리고 기지정된 값을 갖는  $\alpha_1$  및  $\alpha_2$ 는 마진 값이다.

[0083] 수학적 식 4에서 첫번째 항목인  $\|F_{b,a} \circ F_{a,b}\|_2 - \|F_{a,b} \circ F_{b,a}\|_2 + \alpha_1$ 는 오프셋의 합( $F_{a,b} \circ F_{b,a}$ )이 오프셋의 합( $F_{b,a} \circ F_{a,b}$ )과 마진( $\alpha_1$ )의 합보다 작은 경우에 1로 출력되어 신뢰할 수 있음을 의미한다. 그러나 첫번째 항목만으로 신뢰도를 계산하는 경우, 오프셋의 합( $F_{a,b} \circ F_{b,a}$ )과 오프셋의 합( $F_{b,a} \circ F_{a,b}$ )이 마진( $\alpha_1$ )보다 작은 차이를 갖되 둘 다 매우 큰 값을 갖는 경우, 신뢰도를 잘못 판단하게 되는 문제가 발생할 수 있다. 이에 두번째 항목인  $\alpha_2 - \|F_{a,b} \circ F_{b,a}\|_2$ 가 오프셋의 합( $F_{a,b} \circ F_{b,a}$ )이 마진( $\alpha_2$ )보다 작은 경우에 1로 출력되어 신뢰할 수 있음을 의미한다. 즉 수학적 식 4는 오프셋의 합( $F_{a,b} \circ F_{b,a}$ )과 오프셋의 합( $F_{b,a} \circ F_{a,b}$ )이 마진( $\alpha_1$ ) 이내의 차를 갖고, 오프셋의 합( $F_{a,b} \circ F_{b,a}$ )이 마진( $\alpha_2$ )보다 작은 값을 갖는 경우에 신뢰할 수 있음을 나타낸다.

[0084] 그리고 사이클 일관성 손실 측정부(321)는 2개의 프레임의 스테레오 영상 세트의 4개의 영상( $(l_1, r_1)$ ,  $(l_2, r_2)$ )에 대한 순방향 및 역방향 사이클 전이 결과를 계산하므로, 이에 대응하여, 신뢰도 맵 생성부(324) 또한 4개의 영상들 사이에 대한 순방향 및 역방향 신뢰도 맵을 각각 생성할 수 있다.

[0085] 신뢰도 맵 생성부(324)가 a, b, c, d의 순서로 4개의 영상에 대해 a 영상으로부터 b 영상 방향으로의 사이클 전이에 대한 신뢰도 맵을 생성하는 경우, 사이클 신뢰도 맵( $\bar{M}_{a,b(p)}$ )은 수학적 식 5와 같이 계산될 수 있다.

### 수학식 5

$$\begin{aligned} \bar{M}_{a,b}(p) = & M_{a,b}(p) \cdot M_{b,c}(p + F_{a,b}(p)) \cdot \\ & M_{c,d}(p + (F_{a,b} \circ F_{b,c})(p)) \cdot \\ & M_{d,a}(p + (F_{a,b} \circ F_{b,c} \circ F_{c,d})(p)) \end{aligned}$$

[0086]

[0087] 그리고 사이클 일관성 손실 측정부(321)는 신뢰도 맵 생성부(324)에서 획득된 사이클 신뢰도 맵( $\bar{M}_{a,b}(p)$ )에 기반하여, 사이클 손실( $L_c$ )을 수학식 6에 따라 획득할 수 있다.

### 수학식 6

$$\mathcal{L}_c = \mathcal{L}_{l_1, l_2}^c + \mathcal{L}_{l_1, r_1}^c$$

[0088]

[0089] 여기서  $\mathcal{L}_{l_1, l_2}^c$  과  $\mathcal{L}_{l_1, r_1}^c$  는 도5 의 (a) 및 (b)에 나타난 순방향 사이클 손실 및 역방향 사이클 손실을 나타내고, 각각 수학식 7과 같이 계산될 수 있다.

### 수학식 7

$$\mathcal{L}_{a,b}^c = \sum_p \psi(\bar{M}_{a,b}(p) \cdot \bar{F}_{a,b}(p))$$

[0090]

[0091] 수학식 7에서  $\psi()$ 는 절단 함수로서,  $\psi(x) = \min(\|x\|_1, T)$  로서  $\| \cdot \|_1$ 은 L1 norm 함수이며, T는 기지정된 문턱값을 나타낸다. 즉  $\psi(x)$ 는 x의 L1 norm 값과 기지정된 문턱값(T) 중 작은 값을 출력하는 함수로서, 사이클 손실의 이상 출력값을 제거하기 위해 적용된다.

[0092] 한편 사이클 일관성 손실 측정부(321)가 순방향 및 역방향으로 사이클 손실을 측정하여, 사이클 손실이 매우 낮게 나타나더라도 경우에 따라서는 잘못된 사이클 경로를 통해 사이클 손실이 낮게 측정되는 경우가 발생할 수 있다. 즉 4개의 영상 사이를 전이하는 동안 부정확한 대응점을 거치더라도, 최종적으로 초기 픽셀로 전이됨으로써, 사이클 손실이 0 또는 0에 근접한 낮은 값으로 측정되는 경우가 발생할 수 있다.

[0093] 이러한 문제를 보완하기 위해, 복원 손실 측정부(322)는 대응점 사이의 픽셀값 및 그래디언트(gradient) 일관성 오차를 복원 손실로서 측정한다. a 영상과 b 영상 사이의 오프셋( $F_{a,b}$ )이 주어지면, 신뢰도 맵( $M_{a,b}$ )을 적용하여 두 영상 사이의 복원 손실( $L_{a,b}^r$ )은 수학식 8에 따라 획득될 수 있다.

### 수학식 8

$$\begin{aligned} \mathcal{L}_{a,b}^r = & \sum_p M_{a,b}(p) (\|a(p) - b(p + F_{a,b}(p))\|_1) \\ & + \gamma \sum_p M_{a,b}(p) (\|\nabla a(p) - \nabla b(p + F_{a,b}(p))\|_1) \end{aligned}$$

[0094]

[0095] 여기서  $\nabla$  는 그래디언트 연산자이고,  $\gamma$  는 픽셀값과 그래디언트의 균형을 조절하기 위한 밸런스값이다. 수학식 8에서 그래디언트 연산자는 조명의 변화에 강건한 복원 손실을 획득하기 위해 적용되며, 기존에도 옵티컬플로우



추정 시에 널리 이용되고 있는 연산자이다.

[0096] 그리고 수학식 8에서 신뢰도 맵( $M_{a,b}$ )을 적용하는 것은 사이클 일관성 손실 계산에서와 마찬가지로 대응점이 없는 영역에 대한 복원 손실이 반영되지 않도록 하기 위해서이다.

[0097] 도7 및 도8 은 복원 손실의 개념을 설명하기 위한 도면으로, 도7 에서 (a)는 2개의 스테레오 영상 세트( $(l_1, r_1), (l_2, r_2)$ )에서 옵티컬플로우의 순방향 및 역방향 복원 손실을, (b)는 디스패리티의 순방향 및 역방향 복원 손실을 나타내며, 도8 의 (a)와 (b)는 옵티컬플로우와 디스패리티의 순방향 및 역방향 조합 복원 손실을 나타낸다.

[0098] 도7 및 도8 에 나타난 바와 같이, 복원 손실 또한 사이클 일관성 손실과 유사하게 전이 경로를 따라 옵티컬플로우와 디스패리티 및 옵티컬플로우와 디스패리티 조합 각각의 순방향 및 역방향 복원 손실이 모두 고려되어야 한다.

[0099] 도8 에서 옵티컬플로우와 디스패리티 조합에 대한 복원 손실을 t번째 프레임의 좌영상( $l_1$ )과 t+1번째 프레임의 우영상( $r_2$ )으로부터 곧바로 획득하지 않고, t+1번째 프레임의 좌영상( $l_2$ )을 통해 획득하는 것은 오프셋 획득부(310)의 8개의 삼 CNN 중 t번째 프레임의 좌영상( $l_1$ )과 t+1번째 프레임의 우영상( $r_2$ ) 사이의 오프셋을 획득하는 CNN이 존재하지 않기 때문이다. 만일 오프셋 획득부(310)가 8개를 초과하는 개수의 삼 CNN을 포함하는 경우, 옵티컬플로우와 디스패리티 조합에 대한 복원 손실은 도8 과 다르게 t번째 프레임의 좌영상( $l_1$ )과 t+1번째 프레임의 우영상( $r_2$ ) 사이의 오프셋으로부터 직접 획득될 수도 있다.

[0100] 다만 오프셋 획득부(310)의 삼 CNN의 개수의 증가는 오프셋 획득부(310)의 구조를 복잡하게 하여, 학습부(300)를 구성하기 위한 비용이 증가하게 된다. 따라서 본 실시예에서는 효율성을 고려하여, 오프셋 획득부(310)가 8개의 삼 CNN을 구비하는 것으로 가정한다.

[0101] 이에 복원 손실 측정부(322)는 도7 및 도8 에 도시된 2개의 프레임의 스테레오 영상 세트의 4개의 영상( $(l_1, r_1), (l_2, r_2)$ )에 대해 옵티컬플로우와 디스패리티 및 옵티컬플로우와 디스패리티 조합 각각의 순방향 및 역방향 복원 손실을 모두 반영하여 총 재구성 손실을 수학식 9에 따라 획득한다.

### 수학식 9

$$[0102] \mathcal{L}_r = \mathcal{L}_{l_1, l_2}^r + \mathcal{L}_{l_2, l_1}^r + \mathcal{L}_{l_1, r_1}^r + \mathcal{L}_{r_1, l_1}^r + \mathcal{L}_{l_1, r_2}^r + \mathcal{L}_{r_2, l_1}^r$$

[0103] 한편, 평활화 손실 측정부(323)는 옵티컬플로우 및 디스패리티는 영상에서 객체의 경계, 즉 불연속된 부분에 대응하여 급격하게 변화하는 반면, 객체의 내부 영역 즉 연속된 부분에서는 매끄럽게 변화한다는 특성에 따라 평활화 손실을 측정한다.

[0104] 평활화 손실 측정부(323)는 a 영상과 b 영상의 두 영상에 대한 평활화 손실을 수학식 10에 따라 획득할 수 있다.

### 수학식 10

$$[0105] \mathcal{L}_{a,b}^s = \sum_p \|\nabla F_{a,b}(p)\|_1 \exp^{-\beta \|\nabla a(p)\|_1}$$

[0106] 여기서  $\beta$ 는 평활도 대역폭 조절자로서 기지정된 값을 갖는다.

[0107] 다만 연속되는 프레임 간의 차를 나타내는 옵티컬플로우의 오프셋은 서로 다른 시점 사이의 차를 나타내는 디스패리티의 오프셋보다 작은 경우가 빈번하기 때문에, 이후 손실 역전파부(330)가 옵티컬플로우에 대한 평활화 손

실( $L_o$ )과 디스패리티에 대한 평활화 손실( $L_d$ )에 대해 동일한 손실 가중치를 적용하는 것은 바람직하지 않다.

[0108] 이에 평활화 손실 측정부(323)는 옵티컬플로우에 대한 평활화 손실( $L_o$ )과 디스패리티에 대한 평활화 손실( $L_d$ )에 대해 서로 다른 손실 가중치( $w_o$ ,  $w_d$ )를 적용할 수 있도록 구분하여 측정할 수 있다.

[0109] 손실 역전파부(330)는 손실 측정부(320)에서 측정된 사이클 일관성 손실( $L_c$ ), 복원 손실( $L_r$ ) 및 평활화 손실( $L_o$ ,  $L_d$ ) 각각에 대해 수학적 식 11과 같이 기지정된 손실 가중치( $w_c$ ,  $w_r$ ,  $w_o$ ,  $w_d$ )를 적용하고 합하여, 총 손실( $L$ )을 계산한다.

### 수학적 식 11

$$[0110] \quad \mathcal{L} = w_c \mathcal{L}_c + w_r \mathcal{L}_r + w_o \mathcal{L}_o + w_d \mathcal{L}_d$$

[0111] 그리고 손실 역전파부(330)는 계산된 총 손실( $L$ )을 오프셋 획득부(310)의 8개의 삼 CNN으로 역전파함으로써 8개의 삼 CNN의 가중치를 업데이트하는 방식으로 학습시킨다.

[0112] 즉 학습부(300)는 검증 자료가 레이블 되지 않은 일반의 스테레오 영상을 학습 데이터로 이용하여 오프셋 획득부(310)의 8개의 삼 CNN을 비지도 학습을 수행할 수 있다. 여기서 학습부(300)는 총 손실이 기지정된 기준 손실값 이하가 될 때까지 반복 학습시키거나, 기지정된 기준 학습 횟수까지 반복 학습시킬 수 있다.

[0113] 그리고 학습된 8개의 삼 CNN 들의 가중치 중 추정부(200)에 포함된 CNN에 대응하는 CNN의 가중치를 추정부(200)의 CNN으로 전달함으로써, 추정부(200)의 CNN이 직접 학습된 것과 동일한 효과를 나타낼 수 있다.

[0114] 상기에서는 학습부(300)의 오프셋 획득부(310)와 추정부(200)가 별도로 구성되었으나, 경우에 따라서는 학습부(300)의 오프셋 획득부(310)가 추정부(200)으로 이용될 수도 있다. 즉 추정부(200)의 다수의 CNN은 손실 역전파부(330)에서 역전파되는 손실에 기반하여 직접 학습될 수도 있다.

[0115] 도9 는 본 발명의 일 실시예에 따른 옵티컬플로우 및 디스패리티 추정 방법을 나타낸다.

[0116] 도1 내지 도8 을 참조하여, 도9 의 옵티컬플로우 및 디스패리티 추정 방법을 설명하면, 우선 스테레오 영상 획득부(100)가 학습부(300)의 8개의 삼 CNN을 학습시키기 위한 학습 스테레오 영상을 획득하고, 획득된 학습 스테레오 영상에서 연속된 2개 프레임의 스테레오 영상 세트( $(l_1, r_1)$ ,  $(l_2, r_2)$ )를 학습부(300)로 전달한다(S10).

[0117] 여기서 학습 스테레오 영상은 검증 자료가 별도로 레이블된 기존의 학습 데이터가 아닌 단순히 학습 시에 이용되는 스테레오 영상을 의미한다.

[0118] 학습 스테레오 영상이 획득되면, 학습부(300)가 학습 스테레오 영상을 이용하여 추정부(200)의 다수의 CNN에 대한 학습 가중치를 획득하기 위한 학습을 수행한다(S20).

[0119] 학습부(300)의 오프셋 획득부(310)는 추정부(200)의 다수의 CNN과 동일한 구조를 갖는 8개의 삼 CNN을 이용하여 2개 프레임의 스테레오 영상 세트( $(l_1, r_1)$ ,  $(l_2, r_2)$ )로부터 4개의 옵티컬플로우 오프셋( $(F_{l1,l2})$ ,  $(F_{r2,r1})$ ,  $(F_{l2,l1})$ ,  $(F_{r1,r2})$ )과 4개의 디스패리티 오프셋( $(F_{l1,r1})$ ,  $(F_{r2,l2})$ ,  $(F_{r1,l1})$ ,  $(F_{l2,r2})$ )의 총 8개의 오프셋을 획득한다(S21).

[0120] 학습부(300)는 획득된 8개의 오프셋으로부터 손실을 측정하여 8개의 삼 CNN으로 역전파함으로써, 8개의 삼 CNN을 학습시킨다.

[0121] 8개의 삼 CNN을 학습시키기 위해, 우선 손실 측정부(320)의 신뢰도 맵 생성부(324)는 8개의 오프셋을 이용하여 2개 프레임의 스테레오 영상 세트의 4개의 영상( $(l_1, r_1)$ ,  $(l_2, r_2)$ ) 중 서로 다른 기지정된 조합으로 2개 영상 간의 신뢰도 맵을 생성하는 한편, 생성된 신뢰도 맵을 기반으로 사이클 신뢰도 맵을 생성한다(S22).

[0122] 그리고 사이클 일관성 손실 측정부(321)는 8개의 오프셋으로부터 2개 프레임의 스테레오 영상 세트의 4개의 영상( $(l_1, r_1)$ ,  $(l_2, r_2)$ )에 대해 기지정된 순방향 및 역방향으로의 사이클 전이 결과( $\bar{F}_{l1,l2}$ ,  $\bar{F}_{l1,r1}$ )를 획득하고, 획득된 사이클 전이 결과( $\bar{F}_{l1,l2}$ ,  $\bar{F}_{l1,r1}$ )와 생성된 사이클 신뢰도 맵을 기반으로 2개 프레임의 스테레오 영상 세트



의 4개의 영상( $(l_1, r_1)$ ,  $(l_2, r_2)$ )에 대해 기지정된 순방향 및 역방향으로의 사이클 일관성 손실( $L_c$ )을 수학적 식 6 및 7에 따라 계산한다(S23).

[0123] 한편, 복원 손실 측정부(322)는 2개 프레임의 스테레오 영상 세트의 4개의 영상( $(l_1, r_1)$ ,  $(l_2, r_2)$ ) 중 대응점 사이의 픽셀값 차와 그래디언트 차 및 신뢰도 맵을 이용하여 복원 손실( $L_r$ )을 수학적 식 8 및 9에 따라 계산한다(S24). 여기서 복원 손실 측정부(322)는 수학적 식 9에 나타난 바와 같이, 옵티컬플로우의 순방향 및 역방향 복원 손실, 디스패리티의 순방향 및 역방향 복원 손실 및 옵티컬플로우와 디스패리티의 순방향 및 역방향 조합 복원 손실을 합하여 총 복원 손실( $L_r$ )을 계산할 수 있다.

[0124] 그리고 평활화 손실 측정부(323)는 2개 프레임의 스테레오 영상 세트의 4개의 영상( $(l_1, r_1)$ ,  $(l_2, r_2)$ )에서 옵티컬플로우에 따른 2 영상들( $(l_1, l_2)$ ,  $(r_1, r_2)$ )간 평활화 손실( $L_o$ )과 디스패리티에 따른 2 영상들( $(l_1, r_1)$ ,  $(l_2, r_2)$ )간의 평활화 손실( $L_d$ )을 수학적 식 10에 따라 각각 계산한다(S25).

[0125] 손실 역전과부(330)는 손실 측정부(320)에서 측정된 사이클 일관성 손실( $L_c$ ), 복원 손실( $L_r$ ) 및 평활화 손실( $L_o$ ,  $L_d$ ) 각각에 대해 수학적 식 11과 같이 기지정된 손실 가중치( $w_c$ ,  $w_r$ ,  $w_o$ ,  $w_d$ )를 적용하고 합하여, 총 손실( $L$ )을 계산한다(S26).

[0126] 그리고 계산된 총 손실( $L$ )을 오프셋 획득부(310)의 8개의 삼 CNN으로 역전과하여 8개의 삼 CNN의 학습 가중치를 업데이트함으로써 학습시킨다(S27).

[0127] 학습부(300)는 오프셋 획득부(310)의 8개의 삼 CNN의 학습 횟수가 기지정된 기준 학습 횟수 이상인지 판별한다(S28). 만일 기준 학습 횟수 미만이면, 다시 8개의 오프셋을 획득한다(S21). 그러나 기준 학습 횟수 이상이면, 반복을 종료하고, 8개의 삼 CNN의 학습 가중치 중 추정부(200)의 다수의 CNN에 대응하는 학습 가중치들을 대응하는 CNN으로 전달한다(S29). 즉 추정부(200)의 다수의 CNN로 학습된 가중치를 전달하여, 다수의 CNN을 학습된 상태로 전환한다.

[0128] 이후 스테레오 영상 획득부(100)는 옵티컬플로우 및 디스패리티가 획득되어야 하는 스테레오 영상을 획득하고, 획득된 스테레오 영상에서 연속된 2개 프레임씩 스테레오 영상 세트( $(l_1, r_1)$ ,  $(l_2, r_2)$ )를 추정부(200)로 전달한다(S30).

[0129] 학습부(300)의 8개의 삼 CNN과 동일하게 구성되고 학습 가중치를 인가받아 학습된 추정부(200)의 다수의 CNN 중 옵티컬플로우 추정부(210)를 구성하는 2개의 CNN은 2개의 스테레오 영상 세트( $(l_1, r_1)$ ,  $(l_2, r_2)$ )의 t번째 프레임 및 t+1번째 프레임의 좌영상들( $l_1$ ,  $l_2$ ) 사이의 대응점 변이( $F_{l_1, l_2}$ )와 우영상들( $r_1$ ,  $r_2$ ) 사이의 대응점 변이( $F_{r_1, r_2}$ )를 옵티컬플로우로 추정하고, 디스패리티 추정부(220)를 구성하는 적어도 하나의 CNN은 t번째 프레임의 좌영상( $l_1$ )에서 우영상( $r_1$ )으로의 대응점 변이( $F_{l_1, r_1}$ )를 디스패리티로 추정한다(S40). 또한 t+1번째 프레임의 좌영상( $l_2$ )에서 우영상( $r_2$ )으로의 대응점 변이( $F_{l_2, r_2}$ )도 디스패리티로서 함께 추정할 수 있으며, 경우에 따라서는 t번째 프레임 및 t+1번째 프레임 각각의 우영상( $r_1$ ,  $r_2$ )에서 좌영상( $l_1$ ,  $l_2$ )으로의 대응점 변이( $(F_{r_1, l_1})$ ,  $(F_{r_2, l_2})$ )도 디스패리티로 추정할 수 있다.

[0130] 이하에서는 본 실시예에 따른 옵티컬플로우 및 디스패리티 추정 장치 및 방법의 성능을 검토한다.

표 2

Methods	Supervision	Dataset	KITTI 2012		KITTI 2015		Sintel Clean		Sintel Final	
			train	test	train	test	train	test	train	test
FlowNetS-FC [10]	Yes	FC	8.26	-	-	-	4.50	7.42	5.45	8.43
FlowNet2-FC [34]	Yes	FC	4.09	-	10.06	-	2.02	3.96	3.14	6.02
FlowNet2+ft-K [34]	Yes	FC+K	(1.28)	1.80	(2.30)	11.48%	3.43	-	4.66	-
FlowNetS+ft-S [10]	Yes	FC+S	7.52	9.10	-	-	(3.66)	6.96	(4.44)	7.76
DSTFlow-FC [16]	No	FC	16.98	-	24.30	-	6.93	10.40	7.82	11.11
Occlusion-aware-FC [14]	No	FC	12.95	-	21.30	-	5.23	8.02	6.34	9.08
Ours-FT	No	FT	6.70	-	15.59	-	<b>4.85</b>	8.21	<b>5.96</b>	8.76
BackToBasic-K [13]	No	K	11.30	9.90	-	-	-	-	-	-
DSTFlow-K [16]	No	K	10.43	12.40	16.79	39.00%	7.10	10.95	7.95	11.80
UnFlow(CSS)-K [15]	No	K	3.29	-	8.10	-	-	-	7.91	10.22
Occlusion-aware-K [14]	No	K	3.55	<b>4.2</b>	8.88	<b>31.20%</b>	7.41	-	7.92	-
Ours-K	No	K	<b>2.95</b>	<b>4.6</b>	<b>7.65</b>	<b>34.89%</b>	8.58	-	8.54	-
DSTFlow-S [16]	No	S	15.78	-	23.69	-	(6.16)	10.41	(7.38)	11.28
Occlusion-aware+ft-S [14]	No	FC+S	12.90	-	22.60	-	(4.03)	<b>7.95</b>	(5.95)	9.15
Ours+ft-S	No	FT+S	6.44	-	15.02	-	(4.37)	8.08	(5.76)	<b>8.62</b>

[0131]

[0132]

표2 는 FlyingThings3D과 Sintel clean/final 및 KITTI 2012/2015 데이터 셋에 대해 본 실시예에 따른 옵티컬 플로우 추정 성능을 기존의 옵티컬플로우 추정 방식과 비교한 결과를 나타낸다.

[0133]

여기서 이용된 데이터 셋 중 FlyingThings3D는 옵티컬플로우, 디스패리티 및 디스패리티 변화에 대한 검증 자료가 레이블된 스테레오 영상을 제공하며, 21,818개의 레이블과 4,248개의 테스트 스테레오 영상을 포함한 데이터 셋을 제공한다. Sintel는 각각 1064개의 스테레오 영상이 포함된 clean 버전의 데이터 셋과 final 버전의 데이터 셋을 포함한다. 그리고 KITTI(Karlsruhe Institute of Technology and Toyota Technological Institute) 2012과 KITTI 2015은 각각 197 및 200개의 레이블 및 195 및 200개의 테스트 스테레오 영상을 포함한 데이터 셋을 제공한다.

[0134]

학습에 이용된 스테레오 영상은 512 X 256 크기를 가지고, 각 데이터 세트에 대해 1700k, 850k 및 200k회 반복 학습을 수행하였으며, 손실 가중치( $w_c$ ,  $w_r$ ,  $w_o$ ,  $w_d$ )를 각각  $w_c = 1$ ,  $w_r = 1$ ,  $w_o = 13.25$  및  $w_d = 3$ 으로 설정하였다. 그리고  $\alpha_1 = 1$ ,  $\alpha_2 = 20$ ,  $\beta = 1$ ,  $\gamma = 1$ ,  $T = 10$ 으로 설정하였다.

[0135]

다만, KITTI 데이터 셋에 대해서는 조명의 변화를 고려하기 위해,  $w_r = 0.3$  및  $\gamma = 1$ 으로 적용하였다.

[0136]

표2 에서 접미어 -FC는 각 옵티컬플로우 추정 방법이 FlyingChairs 데이터 셋을 이용하여 학습된 경우를 나타내고, -FT는 FlyingThings3D 데이터 셋을 이용하여 학습된 경우를 나타내며, -K는 KITTI 데이터 셋을 이용하여 학습된 경우를 나타낸다. 그리고 -S는 Sintel 데이터 셋을 이용하여 학습된 경우를 나타낸다. 그리고 +ft- 은 두 가지 데이터 집합을 사용하여 미세 조정된 결과를 나타낸다.

[0137]

표2 에서의 모든 결과는 검증 자료(ground truth)와 비교할 때 픽셀 단위의 평균 EPE(endpoint error)를 나타내며, EPE가 3% 또는 5%를 초과할 때, 추정된 옵티컬플로우에 오류가 있는 것으로 간주되는 오류 픽셀의 비율(P1-all)을 %로 나타내었다.

[0138]

표2 를 참조하면, FlyingChairs 데이터 셋으로 비지도 학습된 본 실시예는 동일하게 FlyingChairs 데이터 셋으로 비지도 학습된 기존의 DSTFlow-FC나 Occlusion-aware-FC 보다 KITTI 2012/2015 데이터 셋에 대해 뛰어난 성능을 나타냄을 확인할 수 있다. 특히 Sintel clean/final 데이터 셋에 대해서는 모든 비지도 학습 방식 중 가장 우수한 성능을 나타낸다.

[0139]

그리고 KITTI 2012/2015 데이터 셋으로 비지도 학습된 본 실시예는 KITTI 2012/2015 데이터 셋에 대해 가장 우수한 성능을 나타내며, Sintel clean/final 데이터 셋으로 비지도 학습된 본 실시예는 KITTI 2012/2015 데이터 셋에 대해 지도 학습 방식을 이용하는 FlowNet과 유사한 성능을 나타낸다.

[0140]

한편, 표3 은 FlyingThings3D과 KITTI 2012/2015 데이터 셋에 대해 본 실시예에 따른 디스패리티 추정 성능을 기존의 디스패리티 추정 방식과 비교한 결과를 나타낸다.

표 3

Methods	Supervision	Dataset	KITTI 2012		KITTI 2015		FlyingThings3D	
			train		train		test	
			MAE(px)	D1-all(%)	MAE(px)	log RMSE	MAE(px)	
DispNet [11]	Yes	FT	2.38	-	2.19	-	2.02	
DispNet+ft-K [11]	Yes	FT+K	1.77	-	(0.77)	-	7.14	
Ours-FT	No	FT	2.27	22.91	2.38	0.18	7.35	
Godard <i>et al.</i> [18]	No	K	-	23.81	-	0.18	-	
Godard <i>et al.</i> stereo [18]	No	K	-	9.19	-	0.15	-	
Ours-K	No	K	1.57	12.33	1.65	0.14	8.56	

표3에서는 디스패리티 추정 결과를 검증 자료와 비교한 MAE(mean absolute error), 로그 RMSE(log root mean square error)를 나타내며, D1-all은 MAE가 3% 또는 5% 이상일 때 오류로 간주되는 오류 픽셀의 비율을 나타낸다.

표3을 참조하면, FlyingThings3D 데이터 셋으로 비지도 학습된 본 실시예는 KITTI 2012/2015 데이터 셋에 대해, 동일한 FlyingThings3D 데이터 셋으로 지도 학습된 DispNet 유사한 성능을 제공함을 알 수 있다. 또한 KITTI 2012/2015 데이터 셋으로 비지도 학습된 본 실시예는 KITTI 2012/2015 데이터 셋에 대해 평균 MAE 및 로그 RMSE 측면에서 모든 지도 학습 및 비지도 학습보다 나은 성능을 나타냄을 알 수 있다.

표4는 본 실시예의 디스패리티 추정 성능을 정량적으로 비교한 결과를 나타낸다.

표 4

Methods	Supervision	Data	cap	Abs Rel	Sq Rel	RMSE	RMSE(log)	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
				lower is better			higher is better			
Eigen <i>et al.</i> [35]	Yes	D	0 ~ 80 m	0.215	1.515	7.156	0.270	0.692	0.899	0.967
Kuznetsov <i>et al.</i> [38]	Yes	D & B	0 ~ 80 m	0.113	0.741	4.621	0.189	0.862	0.960	0.986
Luo <i>et al.</i> [39]	Yes	D & B	0 ~ 80 m	0.094	0.626	4.252	0.177	0.891	0.965	0.984
Zhou <i>et al.</i> [40]	No	M	0 ~ 80 m	0.183	1.595	6.709	0.270	0.734	0.902	0.959
Yin <i>et al.</i> [41]	No	M	0 ~ 80 m	0.155	1.296	5.857	0.233	0.793	0.931	0.973
Wang <i>et al.</i> [42]	No	M	0 ~ 80 m	0.151	1.257	5.583	0.228	0.810	0.936	0.974
Godard <i>et al.</i> [18]	No	B	0 ~ 80 m	0.148	1.344	5.297	0.247	0.803	0.922	0.964
Ours-E	No	B	0 ~ 80 m	0.090	0.733	4.363	0.184	0.911	0.961	0.979
Garg <i>et al.</i> [43]	No	B	1 ~ 50 m	0.169	1.080	5.104	0.273	0.740	0.904	0.962
Godard <i>et al.</i> [18]	No	B	1 ~ 50 m	0.108	0.657	3.729	0.194	0.873	0.954	0.979
Ours-E	No	B	1 ~ 50 m	0.085	0.524	3.275	0.172	0.920	0.964	0.981

표4를 참조하면, 본 실시예의 디스패리티 추정 성능은 모든 비지도 학습 방식보다 우수하며, 지도 학습 방식에 근접한 성능을 나타냄을 알 수 있다.

도10 및 도11은 본 실시예에 따른 옵티컬플로우 및 디스패리티 추정 장치 및 방법의 성능을 비교한 결과를 나타낸다.

도10은 본 실시예의 옵티컬플로우 및 디스패리티 추정 성능을 정성적으로 비교한 결과를 나타내며, 도10에서 (a)는 원본 스테레오 영상들을 나타내고, (b)는 검증 자료(ground-truth) 옵티컬 플로우를 나타내며, (c)는 본 실시예의 비지도 학습 방식으로 학습된 경우에 추정되는 옵티컬플로우를 나타낸다. 그리고 (d)는 검증 자료 디스패리티를 나타내고, (e)는 본 실시예의 비지도 학습 방식으로 학습된 경우에 추정되는 디스패리티를 나타낸다.

도10 상단으로부터 첫 6행은 각각 FT로 우선 학습하고, +ft-S로 미세조정 학습된 경우를 나타내고, 이후 6행은 KITTI 2012/2015 데이터 셋으로 비지도 학습되고, KITTI 2012/2015 데이터 셋에 대해 옵티컬플로우와 디스패리티를 추정한 결과를 나타낸다. 하단 3행은 FlyingThings3D 데이터 셋으로 비지도 학습된 경우에 FlyingThings3D 데이터 셋에 대해 옵티컬플로우와 디스패리티를 추정한 결과를 나타낸다.

도10에 도시된 바와 같이, 본 실시예는 검증 자료(ground-truth)에 비견될 수 있는 우수한 성능으로 옵티컬플로우와 디스패리티를 추정할 수 있음을 알 수 있다.

표5는 학습부(300)가 신뢰도 맵, 사이클 일관성 각각 또는 모두를 이용하고, 옵티컬플로우와 디스패리티 각각 또는 모두에 대해 학습하는 경우에 대한 성능 비교를 나타낸다.

표 5

Confidence maps	Cycle consistency	Optical flow	Disparity	FlyingThings3D (test)		Sintel clean (train)		Sintel final (train)	
				Flow	Disparity	Flow	Disparity	Flow	Disparity
		✓	✓	20.70	8.65	6.42	22.13	7.40	24.84
✓		✓	✓	17.00	7.60	5.24	10.62	6.32	13.41
	✓	✓	✓	21.05	8.48	6.44	21.89	7.26	24.81
✓	✓	✓		21.35	-	6.10	-	6.75	-
✓	✓		✓	-	7.81	-	14.38	-	15.55
✓	✓	✓	✓	16.21	7.35	4.85	10.39	5.96	12.13

[0152]

[0153]

표5 를 참조하면, 신뢰도 맵, 사이클 일관성을 모두 이용하고, 옵티컬플로우와 디스패리티를 동시에 학습하는 경우에, FlyingThings3D 및 Sintel clean/final에서 가장 우수한 성능을 나타낼 수 있다.

[0154]

도11 은 학습 방식에 따른 옵티컬플로우와 디스패리티 추정 성능을 정성적으로 나타낸 도면으로, 상단은 옵티컬 플로우 추정 결과를 나타내고, 하단은 디스패리티 추정 결과를 나타낸다. 그리고 (a)는 사이클 일관성 손실 및 신뢰도 맵을 모두 이용하지 않고 학습된 경우를 나타내고, (b)는 사이클 일관성 손실을 이용하지 않는 경우, (c)는 신뢰도 맵을 이용하지 않는 경우, (d)는 옵티컬플로우 및 디스패리티 각각에 대해 개별적으로 학습된 경우, (e)는 본 실시예에 따라 사이클 일관성 손실 및 신뢰도 맵을 이용하여 옵티컬플로우와 디스패리티가 동시 학습된 경우를 나타내며, (f)는 검증 자료를 나타낸다.

[0155]

도11 에 나타난 바와 같이, 본 실시예에 따라 사이클 일관성 손실 및 신뢰도 맵을 이용하여 옵티컬플로우와 디스패리티가 동시 학습된 경우에 검증 자료와 비교한 EPE 및 MAE가 가장 우수하게 나타남을 확인할 수 있다.

[0156]

결과적으로 본 실시예에 따른 옵티컬플로우 및 디스패리티 추정 장치 및 방법은 옵티컬플로우와 디스패리티를 별도로 학습시키지 않아도 되므로, 학습 속도를 향상시킬 수 있을 뿐만 아니라, 개별 학습 방식보다 더욱 우수한 성능을 나타낼 수 있으며, 옵티컬플로우와 디스패리티를 동시에 추정할 수 있도록 한다.

[0157]

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, 광데이터 저장장치 등을 포함할 수 있다.

[0158]

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.

[0159]

따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 청구범위의 기술적 사상에 의해 정해져야 할 것이다.

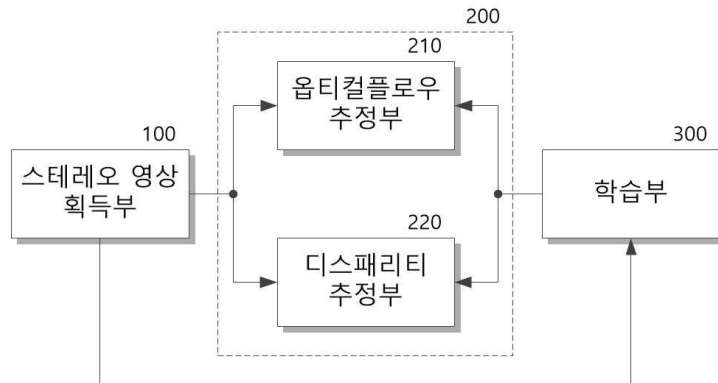
부호의 설명

[0160]

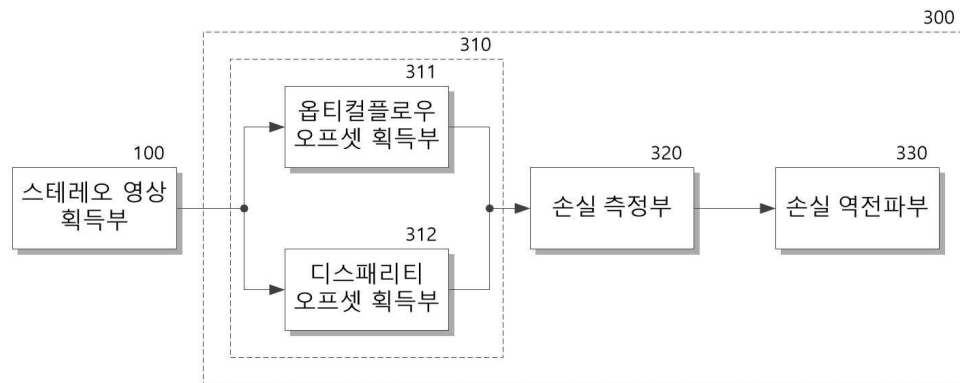
100: 스테레오 영상 획득부 200: 추정부  
 300: 학습부 210: 옵티컬플로우 추정부  
 220: 디스패리티 추정부 310: 오프셋 획득부  
 320: 손실 측정부 330: 손실 역전파부  
 311: 옵티컬플로우 오프셋 획득부 312: 디스패리티 오프셋 획득부  
 321: 사이클 일관성 손실 측정부 322: 복원 손실 측정부  
 323: 평활화 손실 측정부 324: 신뢰도 맵 생성부

도면

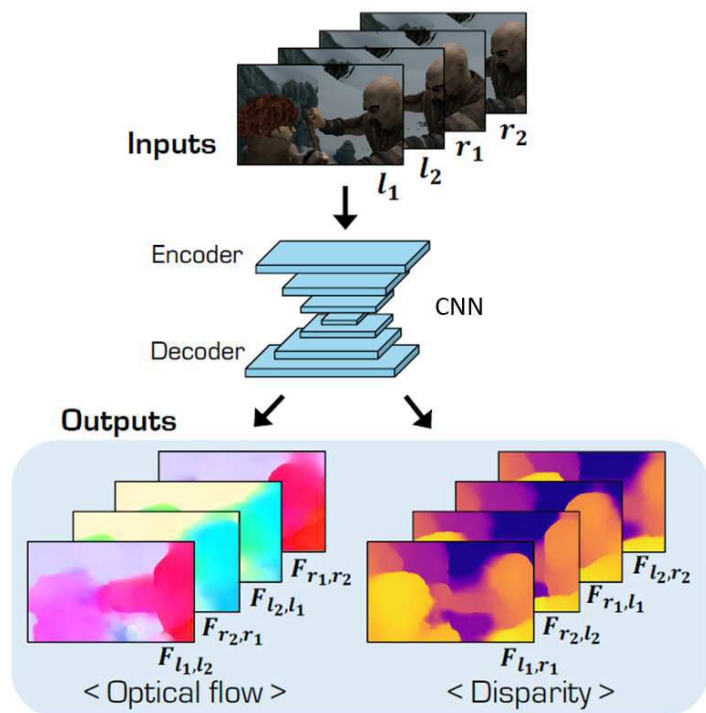
도면1



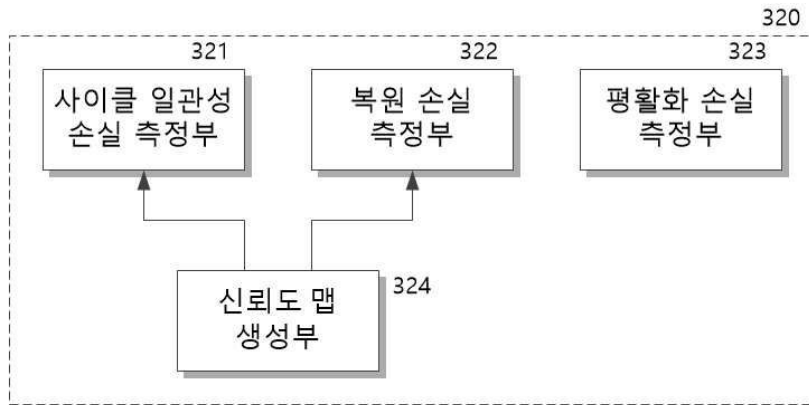
도면2



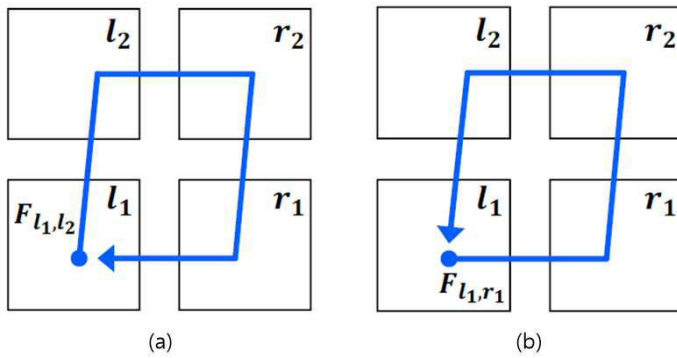
도면3



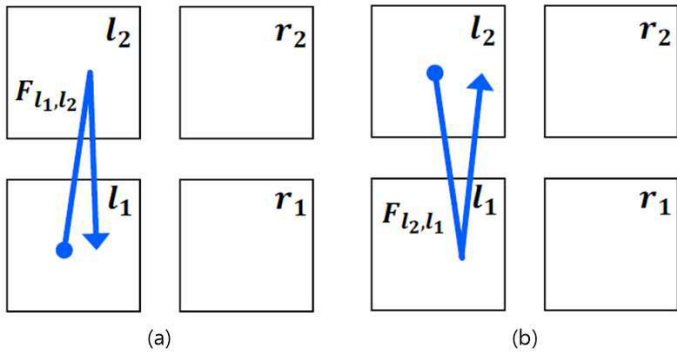
도면4



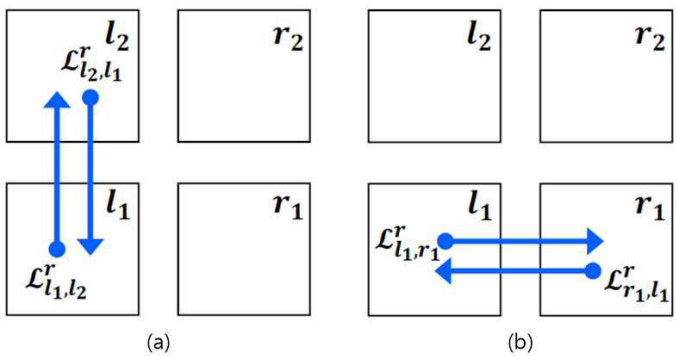
도면5



도면6

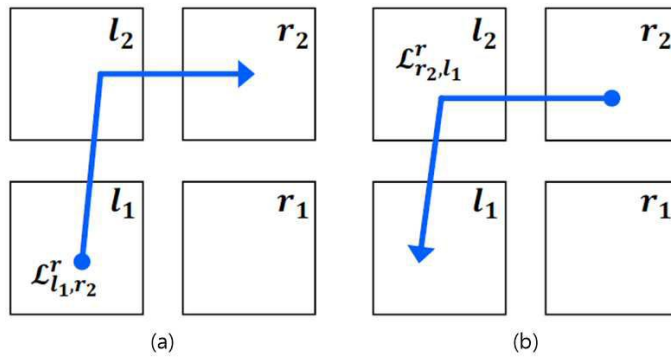


도면7

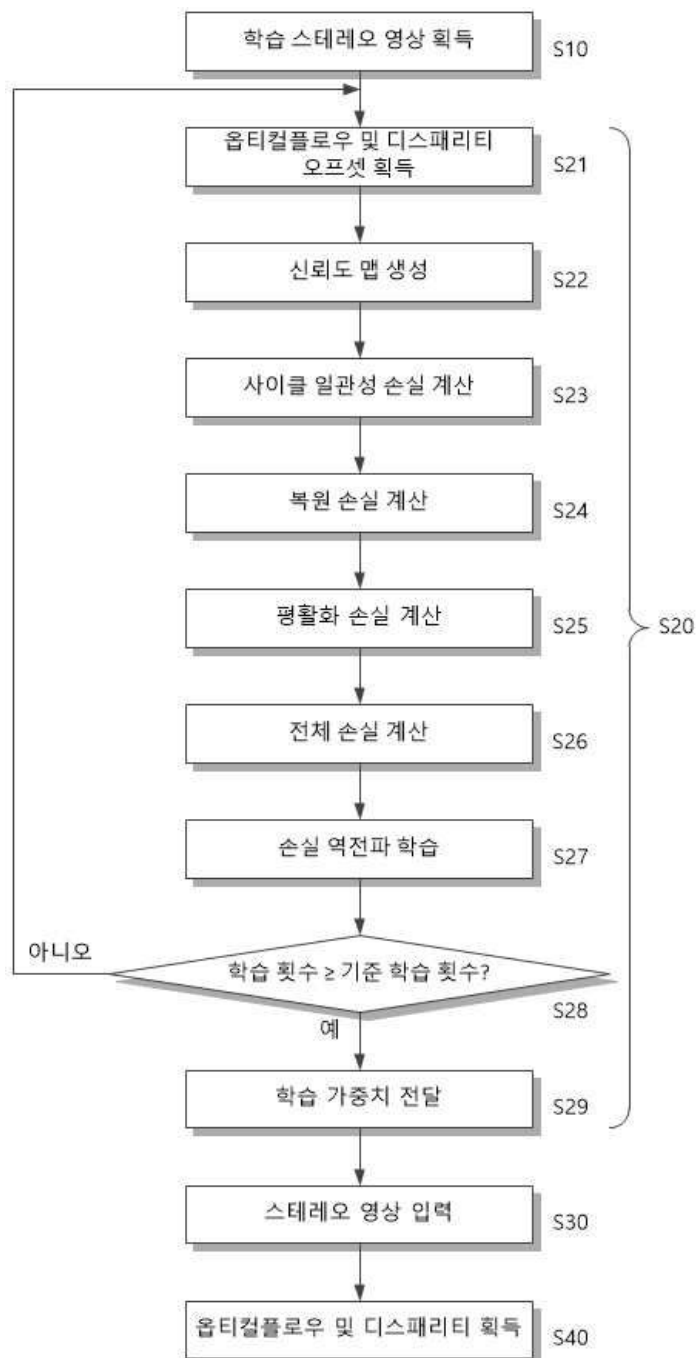




도면8

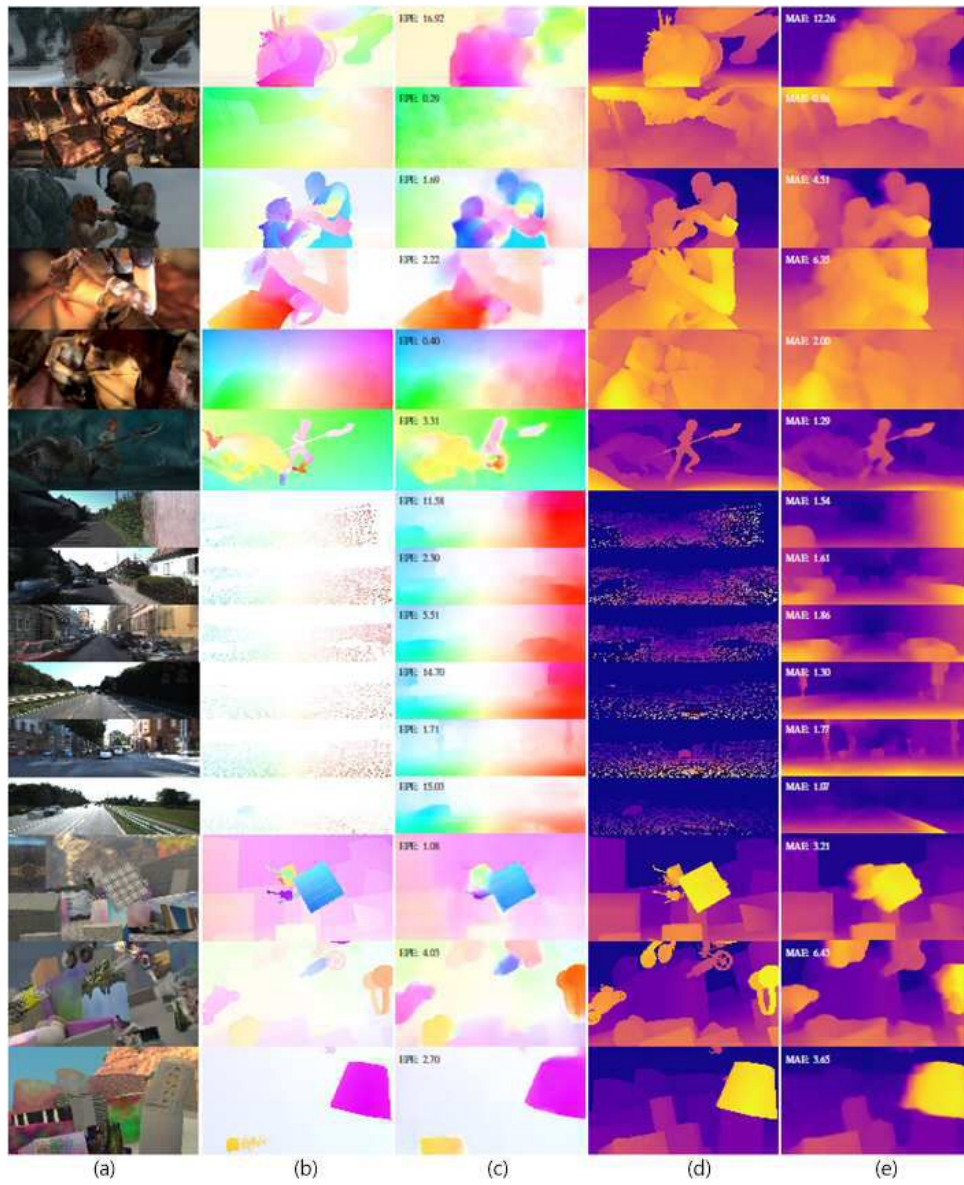


도면9





도면10



도면11

