



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2023년12월04일  
(11) 등록번호 10-2609829  
(24) 등록일자 2023년11월30일

(51) 국제특허분류(Int. Cl.)  
G06N 3/08 (2023.01) G06N 3/04 (2023.01)  
H04N 13/271 (2018.01)  
(52) CPC특허분류  
G06N 3/08 (2023.01)  
G06N 3/04 (2023.01)  
(21) 출원번호 10-2020-0162095  
(22) 출원일자 2020년11월27일  
심사청구일자 2020년11월27일  
(65) 공개번호 10-2022-0074044  
(43) 공개일자 2022년06월03일  
(56) 선행기술조사문헌  
KR1020160010120 A  
KR1020170106325 A  
KR1020180087944 A

(73) 특허권자  
연세대학교 산학협력단  
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)  
(72) 발명자  
손광훈  
서울특별시 서대문구 연세로 50, 연세대학교 제3공학관 C129호(신촌동)  
김선욱  
서울특별시 서대문구 연세로 50, 연세대학교 제3공학관 C129호(신촌동)  
(74) 대리인  
민영준

전체 청구항 수 : 총 18 항

심사관 : 송근배

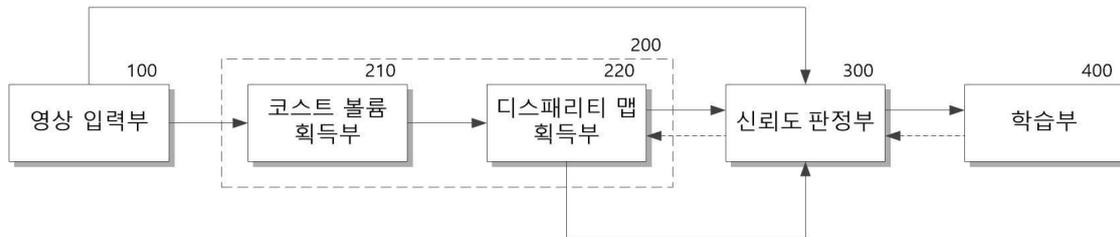
(54) 발명의 명칭 **생성적 적대 신경망을 이용한 스테레오 매칭 신뢰도 추정 장치 및 방법**

(57) 요약

본 발명은 스테레오 영상의 2개의 영상을 인가받아, 2개의 영상 사이에서 기지정된 방식에 따른 픽셀 비교를 통해 각 픽셀별 매칭 코스트를 계산하여 코스트 볼륨을 획득하는 코스트 볼륨 획득부, 미리 학습된 인공 신경망으로 구현되어, 코스트 볼륨으로부터 학습된 방식에 따라 2개의 시점 영상에서 대응하는 픽셀 사이의 거리를 추정

(뒷면에 계속)

대표도 - 도4



하여 디스패리티 맵을 생성하고, 코스트 볼륨의 다수의 매칭 코스트를 학습된 방식에 따라 정제하여 정제 코스트 볼륨을 획득하는 디스패리티 맵 획득부 및 미리 학습된 인공 신경망으로 구현되어, 2개의 영상 중 하나의 영상과 디스패리티 맵 및 정제 코스트 볼륨 각각의 특징을 학습된 방식에 따라 추출하고 추출된 특징을 결합하며, 결합된 특징으로부터 학습된 방식에 따라 디스패리티 맵의 각 픽셀별 신뢰 여부를 판정하여 신뢰도 맵을 획득하는 신뢰도 판정부를 포함하여, 획득된 매칭 코스트에 대한 코스트 집계를 생성하고, 생성된 코스트 집계와 디스패리티 및 원본 스테레오 영상을 기반으로 디스패리티 맵의 신뢰도를 적대적으로 판별하도록 하여, 학습 시에 상호 상승 효과를 유발함으로써, 신뢰도 추정 정확도를 크게 향상시킬 수 있는 스테레오 매칭 신뢰도 추정 장치 및 방법을 제공할 수 있다.

(52) CPC특허분류

*H04N 13/271* (2018.05)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711103201
과제번호	2016-0-00197-005
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신방송연구개발사업
연구과제명	스마트카 다중 센서와 딥러닝을 이용한 초정밀 내추럴 3D 뷰 생성 기술 개발
기 여 율	1/1
과제수행기관명	연세대학교 산학협력단
연구기간	2020.01.01 ~ 2020.12.31
공지예외적용	: 있음

---

**명세서**

**청구범위**

**청구항 1**

스테레오 영상의 2개의 영상을 인가받아, 2개의 영상 사이에서 기지정된 방식에 따른 픽셀 비교를 통해 각 픽셀 별 매칭 코스트를 계산하여 코스트 볼륨을 획득하는 코스트 볼륨 획득부;

미리 학습된 인공 신경망으로 구현되어, 상기 코스트 볼륨으로부터 학습된 방식에 따라 2개의 시점 영상에서 대응하는 픽셀 사이의 거리를 추정하여 디스패리티 맵을 생성하고, 상기 코스트 볼륨의 다수의 매칭 코스트를 학습된 방식에 따라 정제하여 정제 코스트 볼륨을 획득하는 디스패리티 맵 획득부; 및

미리 학습된 인공 신경망으로 구현되어, 상기 2개의 영상 중 하나의 영상과 상기 디스패리티 맵 및 상기 정제 코스트 볼륨 각각의 특징을 학습된 방식에 따라 추출하고 추출된 특징을 결합하며, 결합된 특징으로부터 학습된 방식에 따라 상기 디스패리티 맵의 각 픽셀별 신뢰 여부를 판정하여 신뢰도 맵을 획득하는 신뢰도 판정부를 포함하는 스테레오 매칭 신뢰도 추정 장치.

**청구항 2**

제1항에 있어서, 상기 스테레오 매칭 신뢰도 추정 장치는

학습 시에 결합되어, 학습 스테레오 영상이 상기 코스트 볼륨 획득부에 인가되어 상기 디스패리티 맵 획득부가 상기 디스패리티 맵을 생성하는 과정에서 발생한 디스패리티 손실과 상기 신뢰도 판정부가 디스패리티 맵의 각 픽셀별 신뢰도를 적대적으로 판정하는 과정에서 발생한 신뢰도 손실을 계산하고, 계산된 디스패리티 손실과 신뢰도 손실을 이용하여 계산되는 총 손실을 역전파하여, 인공 신경망으로 구현되는 상기 디스패리티 맵 획득부와 상기 신뢰도 판정부를 학습시키는 학습부를 더 포함하는 스테레오 매칭 신뢰도 추정 장치.

**청구항 3**

제2항에 있어서, 상기 학습부는

학습 스테레오 영상과 함께 제공된 진리 디스패리티 맵과 상기 디스패리티 맵 획득부가 학습 스테레오 영상으로부터 획득한 디스패리티 맵을 기반으로 진리 신뢰도 맵을 획득하고, 진리 신뢰도 맵과 상기 신뢰도 판정부에서 획득된 신뢰도 맵의 픽셀별 신뢰도를 비교하여 신뢰도 맵의 각 픽셀별 신뢰 여부를 양성 또는 음성으로 판별하며, 상기 신뢰도 손실을 판별된 양성 및 음성에 따라 구분하여 계산하는 스테레오 매칭 신뢰도 추정 장치.

**청구항 4**

제3항에 있어서, 상기 학습부는

상기 디스패리티 손실과 양성 및 음성에 따라 구분하여 계산되는 상기 신뢰도 손실 중 음성으로 구분된 신뢰도 손실을 이용하여 상기 총 손실을 계산하는 스테레오 매칭 신뢰도 추정 장치.

**청구항 5**

제4항에 있어서, 상기 학습부는

상기 디스패리티 손실( $L_{disp}(G)$ )을 수학적

$$\begin{aligned} \mathcal{L}_{disp}(G) = & \mathbb{E}_{C \sim p_{data}(C)} [\|G^D(C) - D^*\|_1] \\ & + \mathbb{E}_{C' \sim p_{data}(C')} [\|I^r(G^D(C')) - I^l\|_1] \end{aligned}$$

(여기서  $\mathbb{E}_{C \sim p_{data}(C)}$ 는 스테레오 영상의 2개의 영상 중 하나를 기준 영상( $I^l$ )으로 하여 계산된 코스트 볼륨( $C$ )에서 크로스 엔트로피를 나타내는 에너지 함수이고,  $G^D(C)$ 는 코스트 볼륨( $C$ )에서 추정된 디스패리티 맵이며,  $D^*$ 은 진리 디스패리티 맵이다. 그리고

$C'$  은 기준 영상( $I^1$ )이 아닌 나머지 영상( $I^r$ )을 기준으로 하는 경우 획득되는 와핑 코스트 볼륨을 나타내고,  $G^D(C')$ 은 와핑 코스트 볼륨으로부터 추정되는 와핑 추정 디스페리티 맵을 나타내며,  $I^r G^D(C')$ 은 와핑 추정 디스페리티 맵( $G^D(C')$ )에 따라 와핑된 나머지 영상( $I^r$ )을 나타낸다.  $\mathbb{E}_{C' \sim p_{data}(C')}$  는 와핑 코스트 볼륨( $C'$ )에서의 크로스 엔트로피를 나타내는 에너지 함수이다.)

에 따라 계산하는 스테레오 매칭 신뢰도 추정 장치.

#### 청구항 6

제5항에 있어서, 상기 학습부는

판별된 양성 및 음성에 따른 상기 신뢰도 손실을 구분하여 수학식

$$\begin{aligned} \mathcal{L}_{\text{conf}}(G, F) = & \mathbb{E}_{\hat{C}^t \sim p_{\text{pos}}(C^t), \hat{D}^t \sim p_{\text{pos}}(D^t), I^l \sim p_{\text{pos}}(I^l)} [\log F(\hat{C}^t, \hat{D}^t, I^l)] \\ & + \mathbb{E}_{C \sim p_{\text{neg}}(C), I^l \sim p_{\text{neg}}(I^l)} \left[ \log \left( 1 - F(G(C), I^l) \right) \right] \end{aligned}$$

(여기서  $\hat{C}^t$  와  $\hat{D}^t$  는 각각 t번째 반복 학습에서 양성으로 판별된 코스트 볼륨과 디스페리티 맵을 나타내고,  $\mathbb{E}_{\hat{C}^t \sim p_{\text{pos}}(C^t), \hat{D}^t \sim p_{\text{pos}}(D^t), I^l \sim p_{\text{pos}}(I^l)}$  는 t번째 반복 학습에서 코스트 볼륨( $C^t$ )과 디스페리티 맵( $D^t$ ) 및 컬러 영상( $I^l$ ) 각각의 양성으로 판별된 코스트 볼륨( $\hat{C}^t$ )과 디스페리티 맵( $\hat{D}^t$ ) 및 컬러 영상( $I^l$ ) 픽셀의 크로스 엔트로피를 나타내는 에너지 함수이다. 또한  $\mathbb{E}_{C \sim p_{\text{neg}}(C), I^l \sim p_{\text{neg}}(I^l)}$  는 코스트 볼륨( $C$ )과 컬러 영상( $I^l$ )에서 음성으로 판별된 코스트 볼륨( $C$ )과 컬러 영상( $I^l$ ) 픽셀의 크로스 엔트로피를 나타내는 에너지 함수이고,  $F(G(C))$ 는 학습 과정에서 코스트 볼륨( $C$ )으로부터 추정된 디스페리티 맵( $G(C)$ )에 대해 판정된 신뢰도 맵을 나타낸다.)

에 따라 계산하는 스테레오 매칭 신뢰도 추정 장치.

#### 청구항 7

제6항에 있어서, 상기 학습부는

판별된 양성 및 음성에 따라 구분되어 계산되는 상기 신뢰도 손실 중 음성으로 판별된 신뢰도 손실을

$$\mathcal{L}_{\text{conf}}(G, F) = \mathbb{E}_{C \sim p_{\text{neg}}(C), I^l \sim p_{\text{neg}}(I^l)} \left[ \log \left( 1 - F(G(C), I^l) \right) \right]$$

에 따라 획득하고, 상기 디스페리티 손실( $\mathcal{L}_{\text{disp}}(G)$ )과 상기 음성으로 판별된 신뢰도 손실로부터 상기 총 손실( $\mathcal{L}_{\text{total}}(G, F)$ )을 수학식

$$\mathcal{L}_{\text{total}}(G, F) = \mathcal{L}_{\text{disp}}(G) + \lambda \mathcal{L}_{\text{conf}}(G, F)$$

에 따라 계산하여 역전파하는 스테레오 매칭 신뢰도 추정 장치.

#### 청구항 8

제1항에 있어서, 상기 디스페리티 맵 획득부는

인코더 및 디코더를 구비하는 오토인코더 형식의 인공 신경망으로 구현되어 상기 코스트 볼륨의 매칭 코스트를 학습된 방식에 따라 집계하고, 집계된 코스트 볼륨의 픽셀 각각에 대해 학습된 방식에 따라 디코딩하여 픽셀별 확률 밀도를 추출하는 코스트 볼륨 변환부;

코스트 볼륨의 픽셀별 확률 밀도를 정규화 레이어로 정규화하여 정규 코스트 볼륨을 획득하는 정규화부;

상기 정규 코스트 볼륨의 픽셀별로 정규화된 다수의 매칭 코스트 중 최대가 되는 코스트를 나타내는 디스페리티를 추출하여 상기 디스페리티 맵을 획득하는 디스페리티 추출부; 및

정규 코스트 볼륨에서 학습된 방식에 따라 기설정된 K개의 코스트를 추출 및 풀링하여 상기 정제 코스트 볼륨을 획득하는 코스트 정제부를 포함하는 스테레오 매칭 신뢰도 추정 장치.

**청구항 9**

제8항에 있어서, 상기 신뢰도 판정부는

상기 정제 코스트 볼륨을 인가받아 미리 학습된 방식에 따라 특징을 추출하여 코스트 특징맵을 획득하는 매칭 코스트 특징 추출부;

상기 디스패리티 맵을 인가받아 미리 학습된 방식에 따라 특징을 추출하여 디스패리티 특징맵을 획득하는 디스패리티 특징 추출부;

상기 스테레오 영상의 2개의 영상 중 매칭 코스트 획득 시 기준이 되는 기준 영상을 인가받아 미리 학습된 방식에 따라 특징을 추출하여 컬러 특징맵을 획득하는 컬러 특징 추출부;

상기 코스트 특징맵과 상기 디스패리티 특징맵 및 상기 컬러 특징맵을 인가받아 기설정된 방식으로 결합하여 결합 특징맵을 획득하는 특징 결합부; 및

상기 결합 특징맵을 인가받아 미리 학습된 방식으로 상기 디스패리티 맵의 각 픽셀별 신뢰도를 판정하여 신뢰도 맵을 획득하는 신뢰도 추출부를 포함하는 스테레오 매칭 신뢰도 추정 장치.

**청구항 10**

스테레오 매칭 신뢰도 추정 장치에서 수행되는 스테레오 매칭 신뢰도 추정 방법으로서,

스테레오 영상의 2개의 영상을 인가받아, 2개의 영상 사이에서 기설정된 방식에 따른 픽셀 비교를 통해 각 픽셀별 매칭 코스트를 계산하여 코스트 볼륨을 획득하는 단계;

미리 학습된 인공 신경망을 이용하여, 상기 코스트 볼륨의 다수의 매칭 코스트를 학습된 방식에 따라 정제하여 정제 코스트 볼륨을 획득하고, 상기 코스트 볼륨으로부터 학습된 방식에 따라 2개의 시점 영상에서 대응하는 픽셀 사이의 거리를 추정하여 디스패리티 맵을 생성하는 단계; 및

미리 학습된 인공 신경망을 이용하여, 상기 2개의 영상 중 하나의 영상과 상기 디스패리티 맵 및 상기 정제 코스트 볼륨 각각의 특징을 학습된 방식에 따라 추출하고 추출된 특징을 결합하고, 결합된 특징으로부터 학습된 방식에 따라 상기 디스패리티 맵의 각 픽셀별 신뢰 여부를 판정하여 신뢰도 맵을 획득하는 단계를 포함하는 스테레오 매칭 신뢰도 추정 방법.

**청구항 11**

제10항에 있어서, 상기 스테레오 매칭 신뢰도 추정 방법은

상기 인공 신경망을 학습시키기 위한 학습 단계를 더 포함하고,

상기 학습 단계는 학습 스테레오 영상으로부터 상기 디스패리티 맵을 생성하는 과정에서 발생한 디스패리티 손실을 계산하는 단계;

상기 디스패리티 맵의 각 픽셀별 신뢰도를 적대적으로 판정하는 과정에서 발생한 신뢰도 손실을 계산하는 단계; 및

계산된 디스패리티 손실과 신뢰도 손실을 이용하여 계산되는 총 손실을 역전파하여 상기 인공 신경망을 학습시키는 단계를 포함하는 스테레오 매칭 신뢰도 추정 방법.

**청구항 12**

제11항에 있어서, 상기 디스패리티 손실을 계산하는 단계는

상기 디스패리티 손실( $L_{disp}(G)$ )을 수학적식

$$\begin{aligned} \mathcal{L}_{disp}(G) = & \mathbb{E}_{C \sim p_{data}(C)} [\|G^D(C) - D^*\|_1] \\ & + \mathbb{E}_{C' \sim p_{data}(C')} [\|I^r(G^D(C')) - I^l\|_1] \end{aligned}$$

(여기서  $\mathbb{E}_{C \sim p_{data}(C)}$ 는 스테레오 영상의 2개의 영상 중 하나를 기준 영상( $I^1$ )으로 하여 계산된 코스트 볼륨(C)에서 크로스 엔트로피를 나타내는 에너지 함수이고,  $G^D(C)$ 는 코스트 볼륨(C)에서 추정된 디스패리티 맵이며,  $D^*$ 은 진리 디스패리티 맵이다. 그리고

$C'$ 은 기준 영상( $I^1$ )이 아닌 나머지 영상( $I^r$ )을 기준으로 하는 경우 획득되는 와핑 코스트 볼륨을 나타내고,  $G^D(C')$ 은 와핑 코스트 볼륨으로부터 추정되는 와핑 추정 디스패리티 맵을 나타내며,  $I^r G^D(C')$ 은 와핑 추정 디스패리티 맵( $G^D(C')$ )에 따라 와핑된 나머지 영상( $I^r$ )을 나타낸다.  $\mathbb{E}_{C' \sim p_{data}(C')}$ 는 와핑 코스트 볼륨( $C'$ )에서의 크로스 엔트로피를 나타내는 에너지 함수이다.)

에 따라 계산하는 스테레오 매칭 신뢰도 추정 방법.

### 청구항 13

제12항에 있어서, 상기 신뢰도 손실을 계산하는 단계는

학습 스테레오 영상과 함께 제공된 진리 디스패리티 맵과 학습 스테레오 영상으로부터 생성된 디스패리티 맵을 기반으로 진리 신뢰도 맵을 획득하는 단계;

진리 신뢰도 맵과 획득된 상기 신뢰도 맵의 픽셀별 신뢰도를 비교하여 신뢰도 맵의 각 픽셀별 신뢰 여부를 양성 또는 음성으로 판별하는 단계; 및

상기 신뢰도 손실을 판별된 양성 및 음성에 따라 구분하여 계산하는 단계를 포함하는 스테레오 매칭 신뢰도 추정 방법.

### 청구항 14

제13항에 있어서, 상기 양성 및 음성에 따라 구분하여 계산하는 단계는

판별된 양성 및 음성에 따른 상기 신뢰도 손실을 구분하여 수학식

$$\begin{aligned} \mathcal{L}_{\text{conf}}(G, F) = & \mathbb{E}_{\hat{C}^t \sim p_{\text{pos}}(C^t), \hat{D}^t \sim p_{\text{pos}}(D^t), I^l \sim p_{\text{pos}}(I^l)} [\log F(\hat{C}^t, \hat{D}^t, I^l)] \\ & + \mathbb{E}_{C \sim p_{\text{neg}}(C), I^l \sim p_{\text{neg}}(I^l)} [\log (1 - F(G(C), I^l))] \end{aligned}$$

(여기서  $\hat{C}^t$ 와  $\hat{D}^t$ 는 각각 t번째 반복 학습에서 양성으로 판별된 코스트 볼륨과 디스패리티 맵을 나타내고,  $\mathbb{E}_{\hat{C}^t \sim p_{\text{pos}}(C^t), \hat{D}^t \sim p_{\text{pos}}(D^t), I^l \sim p_{\text{pos}}(I^l)}$ 는 t번째 반복 학습에서 코스트 볼륨( $C^t$ )과 디스패리티 맵( $D^t$ ) 및 컬러 영상( $I^l$ ) 각각의 양성으로 판별된 코스트 볼륨( $\hat{C}^t$ )과 디스패리티 맵( $\hat{D}^t$ ) 및 컬러 영상( $I^l$ ) 픽셀의 크로스 엔트로피를 나타내는 에너지 함수이다. 또한  $\mathbb{E}_{C \sim p_{\text{neg}}(C), I^l \sim p_{\text{neg}}(I^l)}$ 는 코스트 볼륨(C)과 컬러 영상( $I^l$ )에서 음성으로 판별된 코스트 볼륨(C)과 컬러 영상( $I^l$ ) 픽셀의 크로스 엔트로피를 나타내는 에너지 함수이고,  $F(G(C))$ 는 학습 과정에서 코스트 볼륨(C)으로부터 추정된 디스패리티 맵( $G(C)$ )에 대해 판정된 신뢰도 맵을 나타낸다.)

에 따라 계산하는 스테레오 매칭 신뢰도 추정 방법.

### 청구항 15

제14항에 있어서, 상기 학습시키는 단계는

상기 디스패리티 손실과 양성 및 음성에 따라 구분하여 계산되는 상기 신뢰도 손실 중 음성으로 구분된 신뢰도 손실을 이용하여 상기 총 손실을 계산하는 스테레오 매칭 신뢰도 추정 방법.

### 청구항 16

제15항에 있어서, 상기 총 손실을 계산하는 단계는

판별된 양성 및 음성에 따라 구분되어 계산되는 상기 신뢰도 손실 중 음성으로 판별된 신뢰도 손실을

$$\mathcal{L}_{\text{conf}}(G, F) = \mathbb{E}_{C \sim p_{\text{neg}}(C), I^l \sim p_{\text{neg}}(I^l)} \left[ \log \left( 1 - F(G(C), I^l) \right) \right]$$

에 따라 획득하는 단계; 및

상기 총 손실( $\mathcal{L}_{\text{total}}(G, F)$ )을 상기 디스패리티 손실( $\mathcal{L}_{\text{disp}}(G)$ )과 상기 음성으로 판별된 신뢰도 손실로부터 수학적

$$\mathcal{L}_{\text{total}}(G, F) = \mathcal{L}_{\text{disp}}(G) + \lambda \mathcal{L}_{\text{conf}}(G, F)$$

에 따라 계산하여 역전파하는 단계를 포함하는 스테레오 매칭 신뢰도 추정 방법.

#### 청구항 17

제10항에 있어서, 상기 디스패리티 맵을 생성하는 단계는

인코더 및 디코더를 구비하는 오토인코더 형식의 인공 신경망을 이용하여 상기 코스트 볼륨의 매칭 코스트를 학습된 방식에 따라 집계하고, 집계된 코스트 볼륨의 픽셀 각각에 대해 학습된 방식에 따라 디코딩하여 픽셀별 확률 밀도를 추출하는 단계;

코스트 볼륨의 픽셀별 확률 밀도를 정규화 레이어로 정규화하여 정규 코스트 볼륨을 획득하는 단계;

상기 정규 코스트 볼륨의 픽셀별로 정규화된 다수의 매칭 코스트 중 최대가 되는 코스트를 나타내는 디스패리티를 추출하여 상기 디스패리티 맵을 획득하는 단계; 및

정규 코스트 볼륨에서 학습된 방식에 따라 기설정된 K개의 코스트를 추출 및 풀링하여 상기 정제 코스트 볼륨을 획득하는 단계를 포함하는 스테레오 매칭 신뢰도 추정 방법.

#### 청구항 18

제17항에 있어서, 상기 신뢰도 맵을 획득하는 단계는

상기 정제 코스트 볼륨을 인가받아 미리 학습된 방식에 따라 특징을 추출하여 코스트 특징맵을 획득하는 단계;

상기 디스패리티 맵을 인가받아 미리 학습된 방식에 따라 특징을 추출하여 디스패리티 특징맵을 획득하는 단계;

상기 스테레오 영상의 2개의 영상 중 매칭 코스트 획득 시 기준이 되는 기준 영상을 인가받아 미리 학습된 방식에 따라 특징을 추출하여 컬러 특징맵을 획득하는 단계;

상기 코스트 특징맵과 상기 디스패리티 특징맵 및 상기 컬러 특징맵을 인가받아 기설정된 방식으로 결합하여 결합 특징맵을 획득하는 단계; 및

상기 결합 특징맵을 인가받아 미리 학습된 방식으로 상기 디스패리티 맵의 각 픽셀별 신뢰도를 판정하여 신뢰도 맵을 획득하는 단계를 포함하는 스테레오 매칭 신뢰도 추정 방법.

### 발명의 설명

#### 기술 분야

[0001] 본 발명은 스테레오 매칭 신뢰도 추정 장치 및 방법에 관한 것으로, 생성적 적대 신경망을 이용한 스테레오 매칭 신뢰도 추정 장치 및 방법에 관한 것이다.

#### 배경 기술

[0002] 스테레오 매칭(stereo matching)은 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, 정확한 깊이 영상을 획득하기 위하여 다양한 스테레오 매칭 방법이 연구되어 왔다.

[0004] 도 1은 기존의 스테레오 매칭 신뢰도 추정 장치의 개략적 구성을 나타내고, 도 2는 도 1의 진리 신뢰도 맵을 획

득하는 동작을 설명하기 위한 도면이다.

- [0005] 도 1에서 영상 입력부(100)는 서로 다른 2개의 시점 영상( $I^l$ ,  $I^r$ )을 갖는 스테레오 영상을 획득한다. 그리고 스테레오 매칭부(20)는 기지정된 방식에 따라 입력된 스테레오 영상의 깊이 영상에 해당하는 디스패리티 맵(disparity map)을 획득한다.
- [0006] 스테레오 매칭부(20)는 스테레오 영상의 2개의 시점 영상( $I^l$ ,  $I^r$ )에서 각 픽셀 사이의 비유사성을 측정하는 매칭 코스트(matching cost)를 계산하여 코스트 볼륨(C)을 획득하는 코스트 볼륨 획득부(21)와 획득된 코스트 볼륨(C)을 기반으로 스테레오 영상의 2개의 시점 영상에서 서로 대응하는 픽셀을 탐색하여 깊이 영상에 해당하는 디스패리티 맵(disparity map)(D)을 획득한다.
- [0007] 그러나 스테레오 매칭은 도 1에 도시된 바와 같이, 스테레오 영상의 폐색 영역(occluded regions), 질감없는 영역(Texture-less regions), 조명의 변화(illumination variations) 또는 빛의 반사 영역(reflected regions) 등과 같은 영상 자체의 모호성(inherent matching ambiguity)으로 인해 정확한 대응 픽셀을 검출할 수 없어 정확한 디스패리티 맵(D)을 획득하기 어렵다는 한계가 있다.
- [0008] 이에 부정확하게 획득되는 디스패리티 맵(D)에 대한 신뢰도를 향상시키기 위해, 신뢰도 추정부(30)는 기지정된 방식으로 디스패리티 맵의 각 픽셀별 신뢰도를 계산하여 신뢰도 맵(confidence map)을 획득한다. 여기서 신뢰도 맵은 획득된 디스패리티 맵(D)의 각 픽셀에 대해 계산된 신뢰도를 기반으로 부정확한 픽셀을 필터링할 수 있도록 함으로써 더 정확한 디스패리티 맵(D)을 획득할 수 있도록 하기 위해 이용될 수 있다.
- [0009] 한편, 최근에는 딥 러닝(Deep learning) 기법의 발전에 따라 디스패리티 맵과 신뢰도 맵을 인공 신경망(Artificial Neural Network)을 이용하여 획득하는 연구가 활발하게 수행되었다. 이에 최근에는 스테레오 매칭부(20)의 디스패리티 맵 획득부(22)와 신뢰도 추정부(30)를 인공 신경망으로 구현하는 경우가 많다.
- [0010] 도 2는 기존에 인공 신경망으로 구성된 스테레오 매칭 신뢰도 추정 장치의 개략적 구성을 나타낸다.
- [0011] 도 1을 참조하면, 인공 신경망으로 구현되는 스테레오 매칭부(20)의 디스패리티 맵 획득부(22)는 코스트 볼륨 획득부(21)에서 획득된 코스트 볼륨(C)을 인가받아 기지정된 방식으로 코스트 볼륨(C)의 매칭 코스트를 집계하여 집계 코스트 볼륨(G(C))과 추정 디스패리티 맵( $G^D(C)$ )을 추정하여 획득하는 코스트 집계 네트워크(Cost Aggregation Network)라 할 수 있으며, 인공 신경망으로 구현되는 신뢰도 추정부(30)는 디스패리티 맵 획득부(22)에서 추정된 집계 코스트 볼륨(G(C))으로부터 신뢰도 맵(F(G(C)))을 추정하는 신뢰도 추정 네트워크(Certainty Estimation Network)라 할 수 있다.
- [0012] 다만, 이와 같이 디스패리티 맵 획득부(22)와 신뢰도 추정부(30)가 인공 신경망으로 구현되는 경우, 스테레오 매칭 신뢰도 추정 장치는 미리 학습되어야만 이용 가능하다.
- [0013] 도 3은 도 2의 스테레오 매칭 신뢰도 추정 장치의 학습을 설명하기 위한 도면이다.
- [0014] 현재 디스패리티 맵 획득부(22)와 신뢰도 추정부(30)를 학습시키는 방식으로는 학습을 위해 미리 준비된 스테레오 영상( $I^l$ ,  $I^r$ )과 이에 대응하는 진리 디스패리티 맵(Ground-truth disparity map)( $D^*$ )을 포함하는 학습 데이터셋을 이용한 지도 학습(supervised learning) 방식이 일반적이다.
- [0015] 도 2를 참조하면, 기존의 지도 학습 방식에서는 (a)와 같은 스테레오 매칭부(20)에서 획득된 추정 디스패리티 맵( $G^D(C)$ )과 (b)의 진리 디스패리티 맵( $D^*$ ) 사이의 차이를 디스패리티 손실( $L_{disp}(G)$ )로서 기지정된 방식으로 계산하여 역전파함으로써 스테레오 매칭부(20)를 학습시킬 수 있다. 또한 추정 디스패리티 맵( $G^D(C)$ )과 진리 디스패리티 맵( $D^*$ ) 사이의 차이에 따른 진리 신뢰도 맵(Ground-truth confidence map)을 (c)와 같이 획득하고, 획득된 진리 신뢰도 맵과 신뢰도 추정부(30)에서 추정된 추정 신뢰도 맵(F(G(C))) 사이의 차이를 기반으로 신뢰도 손실( $L_{conf}(G,F)$ )을 계산하여 역전파함으로써 신뢰도 추정부(30)를 학습시킬 수 있다.
- [0016] 이와 같은 기존의 학습 방식의 경우, 비록 집계 코스트 볼륨(G(C))을 매개로 하는 점에서는 상호 연관성이 있음에도, 스테레오 매칭부(20)와 신뢰도 추정부(30)가 실질적으로 단순 동시 학습 수준으로 학습이 수행되어, 스테레오 매칭부(20)와 신뢰도 추정부(30)의 상호 보완 학습이 잘 이루어지지 않는다는 한계가 있다. 결과적으로 스테레오 매칭부(20)와 신뢰도 추정부(30)의 학습 성능 향상에 한계가 있다.

**선행기술문헌**

**특허문헌**

[0017] (특허문헌 0001) 한국 등록 특허 제10-1941994호(2019.01.18 등록)

**발명의 내용**

**해결하려는 과제**

[0018] 본 발명의 목적은 생성적 적대 신경망 구조의 생성 네트워크를 기반으로 매칭 코스트를 정제하고, 적대적 네트워크를 기반으로 정제된 매칭 코스트를 기반으로 추정된 신뢰도의 정상 여부를 판별하도록 하여 학습 시에 상호 상승 효과를 유발함으로써, 신뢰도 추정 정확도를 크게 향상시킬 수 있는 스테레오 매칭 신뢰도 추정 장치 및 방법을 제공하는데 있다.

[0019] 본 발명의 다른 목적은 향상된 신뢰도 추정 정확도에 기반하여 정확한 디스패리티 맵을 획득할 수 있도록 하는 스테레오 매칭 신뢰도 추정 장치 및 방법을 제공하는데 있다.

**과제의 해결 수단**

[0020] 상기 목적을 달성하기 위한 본 발명의 일 실시예에 따른 스테레오 매칭 신뢰도 추정 장치는 스테레오 영상의 2개의 영상을 인가받아, 2개의 영상 사이에서 기지정된 방식에 따른 픽셀 비교를 통해 각 픽셀별 매칭 코스트를 계산하여 코스트 볼륨을 획득하는 코스트 볼륨 획득부; 미리 학습된 인공 신경망으로 구현되어, 상기 코스트 볼륨으로부터 학습된 방식에 따라 2개의 시점 영상에서 대응하는 픽셀 사이의 거리를 추정하여 디스패리티 맵을 생성하고, 상기 코스트 볼륨의 다수의 매칭 코스트를 학습된 방식에 따라 정제하여 정제 코스트 볼륨을 획득하는 디스패리티 맵 획득부; 및 미리 학습된 인공 신경망으로 구현되어, 상기 2개의 영상 중 하나의 영상과 상기 디스패리티 맵 및 상기 정제 코스트 볼륨 각각의 특징을 학습된 방식에 따라 추출하고 추출된 특징을 결합하며, 결합된 특징으로부터 학습된 방식에 따라 상기 디스패리티 맵의 각 픽셀별 신뢰 여부를 판정하여 신뢰도 맵을 획득하는 신뢰도 판정부를 포함한다.

[0021] 상기 스테레오 매칭 신뢰도 추정 장치는 학습 시에 결합되어, 학습 스테레오 영상이 상기 코스트 볼륨 획득부에 인가되어 상기 디스패리티 맵 획득부가 상기 디스패리티 맵을 생성하는 과정에서 발생한 디스패리티 손실과 상기 신뢰도 판정부가 디스패리티 맵의 각 픽셀별 신뢰도를 적대적으로 판정하는 과정에서 발생한 신뢰도 손실을 계산하고, 계산된 디스패리티 손실과 신뢰도 손실을 이용하여 계산되는 총 손실을 역전파하여, 인공 신경망으로 구현되는 상기 디스패리티 맵 획득부와 상기 신뢰도 판정부를 학습시키는 학습부를 더 포함할 수 있다.

[0022] 상기 학습부는 학습 스테레오 영상과 함께 제공된 진리 디스패리티 맵과 상기 디스패리티 맵 획득부가 학습 스테레오 영상으로부터 획득한 디스패리티 맵을 기반으로 진리 신뢰도 맵을 획득하고, 진리 신뢰도 맵과 상기 신뢰도 판정부에서 획득된 신뢰도 맵의 픽셀별 신뢰도를 비교하여 신뢰도 맵의 각 픽셀별 신뢰 여부를 양성 또는 음성으로 판별하며, 상기 신뢰도 손실을 판별된 양성 및 음성에 따라 구분하여 계산할 수 있다.

[0023] 상기 학습부는 상기 디스패리티 손실과 양성 및 음성에 따라 구분하여 계산되는 상기 신뢰도 손실 중 음성으로 구분된 신뢰도 손실을 이용하여 상기 총 손실을 계산할 수 있다.

[0024] 상기 디스패리티 맵 획득부는 인코더 및 디코더를 구비하는 오토인코더 형식의 인공 신경망으로 구현되어 상기 코스트 볼륨의 매칭 코스트를 학습된 방식에 따라 집계하고, 집계된 코스트 볼륨의 픽셀 각각에 대해 학습된 방식에 따라 디코딩하여 픽셀별 확률 밀도를 추출하는 코스트 볼륨 변환부; 코스트 볼륨의 픽셀별 확률 밀도를 정규화 레이어로 정규화하여 정규 코스트 볼륨을 획득하는 정규화부; 상기 정규 코스트 볼륨의 픽셀별로 정규화된 다수의 매칭 코스트 중 최대가 되는 코스트를 나타내는 디스패리티를 추출하여 상기 디스패리티 맵을 획득하는 디스패리티 추출부; 및 정규 코스트 볼륨에서 학습된 방식에 따라 기설정된 K개의 코스트를 추출 및 풀링하여 상기 정제 코스트 볼륨을 획득하는 코스트 정제부를 포함할 수 있다.

[0025] 상기 신뢰도 판정부는 상기 정제 코스트 볼륨을 인가받아 미리 학습된 방식에 따라 특징을 추출하여 코스트 특징맵을 획득하는 매칭 코스트 특징 추출부; 상기 디스패리티 맵을 인가받아 미리 학습된 방식에 따라 특징을 추출하여 디스패리티 특징맵을 획득하는 디스패리티 특징 추출부; 상기 스테레오 영상의 2개의 영상 중 매칭 코스트

트 획득 시 기준이 되는 기준 영상을 인가받아 미리 학습된 방식에 따라 특징을 추출하여 컬러 특징맵을 획득하는 컬러 특징 추출부; 상기 코스트 특징맵과 상기 디스패리티 특징맵 및 상기 컬러 특징맵을 인가받아 기지정된 방식으로 결합하여 결합 특징맵을 획득하는 특징 결합부; 및 상기 결합 특징맵을 인가받아 미리 학습된 방식으로 상기 디스패리티 맵의 각 픽셀별 신뢰도를 판정하여 신뢰도 맵을 획득하는 신뢰도 추출부를 포함할 수 있다.

[0026] 상기 목적을 달성하기 위한 본 발명의 다른 실시예에 따른 스테레오 매칭 신뢰도 추정 방법은 스테레오 영상의 2개의 영상을 인가받아, 2개의 영상 사이에서 기지정된 방식에 따른 픽셀 비교를 통해 각 픽셀별 매칭 코스트를 계산하여 코스트 볼륨을 획득하는 단계; 미리 학습된 인공 신경망을 이용하여, 상기 코스트 볼륨의 다수의 매칭 코스트를 학습된 방식에 따라 정제하여 정제 코스트 볼륨을 획득하고, 상기 코스트 볼륨으로부터 학습된 방식에 따라 2개의 시점 영상에서 대응하는 픽셀 사이의 거리를 추정하여 디스패리티 맵을 생성하는 단계; 및 미리 학습된 인공 신경망을 이용하여, 상기 2개의 영상 중 하나의 영상과 상기 디스패리티 맵 및 상기 정제 코스트 볼륨 각각의 특징을 학습된 방식에 따라 추출하고 추출된 특징을 결합하고, 결합된 특징으로부터 학습된 방식에 따라 상기 디스패리티 맵의 각 픽셀별 신뢰 여부를 판정하여 신뢰도 맵을 획득하는 단계를 포함한다.

**발명의 효과**

[0027] 따라서, 본 발명의 실시예에 따른 스테레오 매칭 신뢰도 추정 장치 및 방법은 스테레오 매칭부에서 획득된 매칭 코스트에 대한 코스트 집계를 생성하고, 생성된 코스트 집계와 디스패리티 및 원본 스테레오 영상을 기반으로 디스패리티 맵의 신뢰도를 적대적으로 판별하도록 하여, 학습 시에 상호 상승 효과를 유발함으로써, 신뢰도 추정 정확도를 크게 향상시킬 수 있다. 또한 향상된 신뢰도 추정 정확도에 기반하여 정확한 디스패리티 맵을 획득할 수 있도록 한다.

**도면의 간단한 설명**

[0028] 도 1은 기존의 스테레오 매칭 신뢰도 추정 장치의 개략적 구성을 나타낸다.  
 도 2는 기존에 인공 신경망으로 구성된 스테레오 매칭 신뢰도 추정 장치의 개략적 구성을 나타낸다.  
 도 3은 도 2의 스테레오 매칭 신뢰도 추정 장치의 학습을 설명하기 위한 도면이다.  
 도 4는 본 발명의 일 실시예에 따른 스테레오 매칭 신뢰도 추정 장치의 개략적 구성을 나타낸다.  
 도 5는 도 4의 코스트 집계 생성부의 상세 구현 예를 나타낸다.  
 도 6은 도 4의 신뢰도 추정 판별부의 상세 구현 예를 나타낸다.  
 도 7은 인공 신경망으로 구성된 도 4의 스테레오 매칭 신뢰도 추정 장치의 개략적 구성을 나타낸다.  
 도 8은 본 실시예에 따른 스테레오 매칭 신뢰도 추정 장치의 성능을 비교한 결과를 나타낸다.  
 도 9는 본 발명의 일 실시예에 따른 스테레오 매칭 신뢰도 추정 방법을 나타낸다.

**발명을 실시하기 위한 구체적인 내용**

[0029] 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.

[0030] 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.

[0031] 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 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.

[0032] 도 4는 본 발명의 일 실시예에 따른 스테레오 매칭 신뢰도 추정 장치의 개략적 구성을 나타내고, 도 5는 도 4의 코스트 집계 생성부의 상세 구현 예를 나타내며, 도 6은 도 4의 신뢰도 추정 판별부의 상세 구현 예를 나타낸다. 그리고 도 7은 인공 신경망으로 구성된 도 4의 스테레오 매칭 신뢰도 추정 장치의 개략적 구성을 나

타낸다.

- [0033] 도 4를 참조하면, 본 실시예에 따른 스테레오 매칭 신뢰도 추정 장치는 영상 입력부(100), 스테레오 매칭부(200), 신뢰도 판정부(300) 및 학습부(400)를 포함할 수 있다. 즉 본 실시예에 따른 스테레오 매칭 신뢰도 추정 장치 또한 기본적으로, 도 1에 도시된 기존의 스테레오 매칭 추정 장치와 유사한 구성을 갖는다.
- [0034] 영상 입력부(100)는 스테레오 매칭을 수행해야할 스테레오 영상을 획득한다. 여기서 스테레오 영상은 스테레오 카메라에서 획득될 수 있는 영상으로 서로 다른 시점을 갖는 2개의 시점 영상으로 구성될 수 있다. 스테레오 영상은 스테레오 카메라의 구조에 따라 상하 영상 좌우 영상 등으로 획득될 수 있으나, 여기서는 일 예로 도 1에 도시된 바와 같이, 좌 영상(Left image)( $I^l$ )과 우 영상(Right image)( $I^r$ )으로 구성되는 스테레오 영상을 획득하는 것으로 가정한다.
- [0035] 스테레오 매칭부(200)는 영상 입력부(100)에서 획득된 스테레오 영상의 2개의 시점 영상 사이의 차이를 기반으로 디스패리티 맵(D)을 획득한다.
- [0036] 스테레오 매칭부(200)는 코스트 볼륨 획득부(210) 및 디스패리티 맵 획득부(220)를 포함할 수 있다.
- [0037] 코스트 볼륨 획득부(210)는 영상 입력부(100)에서 획득된 스테레오 영상인 좌 영상( $I^l$ )과 우 영상( $I^r$ )을 인가받고, 인가된 좌 영상( $I^l$ )과 우 영상( $I^r$ ) 중 하나를 기준(여기서는 일 예로 좌 영상( $I^l$ ))으로 각 픽셀별로 나머지 영상(여기서는 우 영상( $I^r$ ))의 픽셀들과의 매칭 코스트(matching cost)를 기지정된 방식으로 획득하여, 매칭 코스트의 집합인 코스트 볼륨(cost volume)(C)을 획득한다.
- [0038] 이때 코스트 볼륨 획득부(210)는 스테레오 영상이 좌우 영상으로 구성된 경우, 기준이 되는 좌 영상( $I^l$ )의 특정 위치(x, y)에 기지정된 크기의 윈도우를 설정하고, 나머지 영상인 우 영상( $I^r$ )에도 동일한 위치에 동일한 크기의 윈도우를 설정한 후, 윈도우를 행 방향(x축 방향)으로 이동시키면서, 윈도우의 이동 거리(d)별로 좌 영상( $I^l$ ) 및 우 영상( $I^r$ )의 윈도우에 포함된 각 픽셀 값을 비교하여 매칭 코스트를 계산할 수 있다.
- [0039] 코스트 볼륨 획득부(210)는 기준이 되는 영상( $I^l$ )의 각 픽셀( $i^l_{x,y}$ )과 비교되는 영상( $I^r$ )의 각 거리(d)에서의 비교 대상 픽셀( $i^r_{(x-d),y}$ ) 사이의 차이로 매칭 코스트( $C_{i,d}$ )를 계산할 수 있으며, 각 픽셀에서 가능한 모든 거리(d)에 대해 획득된 매칭 코스트( $C_{i,d}$ )의 집합을 코스트 볼륨(C)으로 획득할 수 있다.
- [0040] 디스패리티 맵 획득부(220)는 인공 신경망으로 구현되어 매칭 비용 계산부(130)에서 코스트 볼륨(C)이 인가되면, 미리 학습된 방식에 따라 코스트 볼륨(C)으로부터 디스패리티 맵(D)을 추정하여 획득한다. 이때, 도 5에 도시된 바와 같은 인공 신경망으로 구현되는 디스패리티 맵 획득부(220)는 코스트 볼륨 변환부(221), 정규화부(222), 디스패리티 추출부(223) 및 코스트 정제부(224)로 구성될 수 있다.
- [0041] 코스트 볼륨 변환부(221)는 학습에 의해 업데이트되는 가중치( $W_c$ )를 갖는 인코더 및 디코더를 구비하는 오토인코더(auto-encode) 형식의 잔차 컨볼루션 모듈(Residual Convolutional Module)을 이용하여, 코스트 볼륨(C)의 매칭 코스트를 학습된 방식에 따라 집계하고, 집계된 코스트 볼륨(C)의 모든 픽셀( $i = [i_x, i_y]^T$ ) 각각에 대해 학습된 방식에 따라 디코딩하여 각 픽셀별 코스트 볼륨( $C_i$ )에 대한 확률 밀도를 추출한다.
- [0042] 그리고 정규화부(222)는 코스트 볼륨 변환부(221)에서 추출된 각 픽셀별 코스트 볼륨( $C_i$ )에 대한 확률 밀도를 정규화 레이어로 정규화하여 정규 코스트 볼륨( $C_n$ )을 획득한다. 여기서 정규 코스트 볼륨( $C_n$ )은 매칭 확률 볼륨(Matching Probability Volume: MPV)이라고도 할 수 있다.
- [0043] 디스패리티 추출부(223)는 획득된 정규 코스트 볼륨( $C_n$ )을 기반으로 디스패리티 맵(D)을 추정하여 생성한다. 디스패리티 추출부(223)는 소프트 아그맥스 레이어(Soft-Argmax Layer)를 포함하여 정규 코스트 볼륨( $C_n$ )의 각 픽셀별로 정규화된 다수의 매칭 코스트 중 최대가 되는 코스트를 나타내는 디스패리티(d)를 추출함으로써 디스패리티 맵(D)을 획득한다.
- [0044] 한편, 본 실시예에서 코스트 정제부(224)는 정규 코스트 볼륨( $C_n$ )에서 학습된 방식에 따라 기설정된 K개의 코스트

트를 추출 및 풀링하여 정제 코스트 볼륨( $C_r$ )을 더 획득할 수 있다. 여기서 정제 코스트 볼륨( $C_r$ )은 신뢰도 판정부(300)가 획득된 디스패리티 맵(D)의 신뢰도를 판정할 수 있도록 하기 위해 획득되는 코스트 볼륨이다.

[0045] 신뢰도 판정부(300)는 스테레오 매칭부(200)에서 획득된 디스패리티 맵(D)의 각 픽셀별 신뢰도를 판정하여 신뢰도 맵(Q)을 획득한다. 신뢰도 판정부(300)는 스테레오 매칭부(200)의 디스패리티 맵 획득부(220)에서 획득된 디스패리티 맵(D)과 정제 코스트 볼륨( $C_r$ ) 및 스테레오 영상의 2개의 영상( $I^l, I^r$ ) 중 매칭 코스트 계산 신에 기준이 되는 하나의 영상( $I^l$ )을 인가받고, 인가된 디스패리티 맵(D)과 정제 코스트 볼륨( $C_r$ ) 및 기준 영상( $I^l$ )을 기반으로 디스패리티 맵(D)의 각 픽셀에 대한 신뢰도를 판정하여 신뢰도 맵(Q)을 획득한다.

[0046] 도 6을 참조하면, 신뢰도 판정부(300)는 매칭 코스트 특징 추출부(310), 디스패리티 특징 추출부(320), 컬러 특징 추출부(330), 특징 결합부(340) 및 신뢰도 추출부(350)를 포함할 수 있다.

[0047] 매칭 코스트 특징 추출부(310)는 디스패리티 맵 획득부(220)에서 획득된 정제 코스트 볼륨( $C_r$ )을 인가받고 미리 학습된 방식에 따라 정제 코스트 볼륨( $C_r$ )의 특징을 추출하여 코스트 특징맵을 획득한다.

[0048] 디스패리티 특징 추출부(320)는 디스패리티 맵 획득부(220)에서 획득된 디스패리티 맵(D)을 인가받아 미리 학습된 방식에 따라 특징을 추출하여 디스패리티 특징맵을 획득한다.

[0049] 한편, 컬러 특징 추출부(330)는 스테레오 영상의 2개의 시점 영상 중 매칭 코스트 획득 시 기준이 되는 기준 영상( $I_1$ )을 인가받는다. 여기서 스테레오 영상의 2개의 시점 영상은 컬러 영상이므로, 컬러 특징 추출부(330)는 컬러 영상인 기준 영상( $I_1$ )을 인가받아 미리 학습된 방식에 따라 특징을 추출하여 컬러 특징맵을 획득한다.

[0050] 본 실시예에서 신뢰도 판정부(300)가 디스패리티 맵 획득부(220)에서 획득된 정제 코스트 볼륨( $C_r$ )과 디스패리티 맵(D) 뿐만 아니라, 기준 영상( $I_1$ )을 함께 인가받아 특징을 추출하는 것은 스테레오 매칭의 대상이 되는 원본 영상을 신뢰도 판별 시에 함께 이용하여 더욱 정확하게 디스패리티 맵(D)의 신뢰도를 판별할 수 있도록 하기 위함이다.

[0051] 여기서 매칭 코스트 특징 추출부(310), 디스패리티 특징 추출부(320) 및 컬러 특징 추출부(330)를 통합하여 특징 추출부라 할 수 있다.

[0052] 특징 결합부(340)는 매칭 코스트 특징 추출부(310)와 디스패리티 특징 추출부(320) 및 컬러 특징 추출부(330) 각각에서 획득된 코스트 특징맵과 디스패리티 특징맵 및 컬러 특징맵을 인가받아 결합(Concatenation)하여 결합 특징맵을 획득한다. 특징 결합부(340)는 코스트 특징맵과 디스패리티 특징맵 및 컬러 특징맵을 결합하는 맵 결합부(341)와 코스트 특징맵과 디스패리티 특징맵 및 컬러 특징맵의 가중치를 설정하는 가중치 설정부(342) 및 가중치 설정부(342)에서 설정된 가중치를 결합된 코스트 특징맵과 디스패리티 특징맵 및 컬러 특징맵 각각에 적용하는 가중치 적용부(343)를 포함할 수 있다.

[0053] 신뢰도 추출부(350)는 코스트 특징맵과 디스패리티 특징맵 및 컬러 특징맵에 가중치가 적용되어 결합된 결합 특징맵을 인가받고, 미리 학습된 방식에 따라 디스패리티 맵(D)의 각 픽셀별 신뢰도를 판정하여 신뢰도 맵(Q)을 획득한다.

[0054] 한편, 도 5 및 도 6에 도시된 바와 같이, 본 실시예에서 디스패리티 맵 획득부(220)와 신뢰도 판정부(300)는 인공 신경망으로 구현될 수 있으며, 특히 디스패리티 맵 획득부(220)와 신뢰도 판정부(300)가 생성적 적대 신경망 구조로 학습될 수 있다. 이에 도 7에 도시된 바와 같이, 디스패리티 맵 획득부(220)는 디스패리티 맵(D)을 생성하는 생성자(Generator)로서 생성적 코스트 집계 네트워크(Generative Cost Aggregation Network)라 할 수 있으며, 신뢰도 판정부(300)는 생성자인 디스패리티 맵 획득부(220)에서 생성된 디스패리티 맵(D)의 신뢰 여부를 판정하여 신뢰도 맵(Q)을 획득하는 판정자(Discriminator)로서 적대적 신뢰도 추정 네트워크(Adversarial Confidence Estimation Network)라 할 수 있다.

[0055] 생성적 적대 신경망으로 이루어지는 디스패리티 맵 획득부(220)와 신뢰도 판정부(300)의 학습 과정 동안 디스패리티 맵 획득부(220)는 학습되는 네트워크 함수(G)에 따라 인가된 코스트 볼륨(C)으로부터 추정 디스패리티 맵( $G^D(C)$ )과 추정 정제 코스트 볼륨( $G^C(C)$ )을 추정하여 생성한다. 여기서 추정 디스패리티 맵( $G^D(C)$ )과 추정 정제 코스트 볼륨( $G^C(C)$ )을 디스패리티 맵(D)과 정제 코스트 볼륨( $C_r$ )과 구분하여 표시한 것은 학습 과정에서 생성자

인 디스패리티 맵 획득부(220)에서 생성되는 추정 디스패리티 맵( $G^D(C)$ )과 추정 정제 코스트 볼륨( $G^C(C)$ )은 부정확하게 생성된 결과이기 때문이다. 그리고 생성적 코스트 집계 네트워크 인 디스패리티 맵 획득부(220)에서 생성되는 추정 디스패리티 맵( $G^D(C)$ )과 추정 정제 코스트 볼륨( $G^C(C)$ )을 통합하여 생성 네트워크 출력( $G(C) = \{G^D(C), G^C(C)\}$ )이라 할 수 있다.

[0056] 한편, 학습 과정 동안 신뢰도 판정부(300)는 학습되는 네트워크 함수(F)에 따라 기준 영상( $I^1$ )을 참조하여, 생성 네트워크 출력( $G(C)$ )의 추정 디스패리티 맵( $G^D(C)$ )의 각 픽셀에 대한 신뢰 여부를 판정하여 신뢰도 맵( $F(G(C), I^1)$ )을 획득한다. 신뢰도 판정부(300)는 추정 디스패리티 맵( $G^D(C)$ )에 대한 각 픽셀별 판정 결과를 기존의 신뢰도 맵과 마찬가지로 참 또는 거짓의 1과 0으로 표현하여 신뢰도 맵( $F(G(C), I^1)$ )을 획득할 수 있다. 즉 추정 디스패리티 맵( $G^D(C)$ )의 픽셀 값이 신뢰할 수 있는 것으로 판정되면 1의 값을 갖고, 신뢰할 수 없는 것으로 판정되면 0을 값을 갖도록 각 픽셀값을 설정하여 신뢰도 맵( $F(G(C), I^1)$ )을 획득할 수 있다. 다만 신뢰도 판정부(300) 또한 디스패리티 맵 획득부(220)와 마찬가지로 학습 과정에 있으므로, 획득되는 신뢰도 맵( $F(G(C), I^1)$ )은 학습 완료 이후의 신뢰도 맵(Q)과 구분되어 추정 신뢰도 맵이라 할 수 있다.

[0057] 학습부(400)는 디스패리티 맵 획득부(220)에서 생성되는 추정 디스패리티 맵( $G^D(C)$ )과 도 3의 (b)와 같이 미리 획득된 진리 디스패리티 맵( $D^*$ ) 사이의 차이를 디스패리티 손실( $L_{disp}(G)$ )로서 계산한다.

[0058] 여기서 디스패리티 손실( $L_{disp}(G)$ )은 수학식 1과 같이 계산될 수 있다.

**수학식 1**

$$L_{disp}(G) = \mathbb{E}_{C \sim p_{data}(C)} [\|G^D(C) - D^*\|_1] + \mathbb{E}_{C' \sim p_{data}(C')} [\|I^r(G^D(C')) - I^1\|_1]$$

[0059] 여기서  $\mathbb{E}_{C \sim p_{data}(C)}$ 는 코스트 볼륨(C)에서 크로스 엔트로피(cross entropy)를 나타내는 에너지 함수이고, C'은 스테레오 영상에서 설정된 기준 영상( $I^1$ )이 아닌 나머지 영상( $I^r$ )을 기준으로 하는 경우 획득되는 와핑 코스트 볼륨을 나타내며,  $G^D(C')$ 은 와핑 코스트 볼륨으로부터 추정되는 와핑 추정 디스패리티 맵을 나타내며,  $I^r G^D(C')$ 은 와핑 추정 디스패리티 맵( $G^D(C')$ )에 따라 와핑된 나머지 영상( $I^r$ )을 나타낸다. 그리고  $\mathbb{E}_{C' \sim p_{data}(C')}$ 는 와핑 코스트 볼륨(C')에서의 크로스 엔트로피를 나타내는 에너지 함수이다.

[0061] 즉 수학식 1에 따르면, 디스패리티 손실( $L_{disp}(G)$ )은 추정 디스패리티 맵( $G^D(C)$ )과 진리 디스패리티 맵( $D^*$ ) 사이의 차이에 따른 크로스 엔트로피와 와핑된 나머지 영상( $I^r G^D(C')$ )과 기준 영상 사이의 차이에 따른 크로스 엔트로피의 합으로 계산될 수 있다.

[0062] 또한 학습부(400)는 신뢰도 판정부(300)에서 추정된 추정 신뢰도 맵( $F(G(C), I^1)$ )을 인가받아 도 3의 (c)와 같이 미리 획득된 진리 신뢰도 맵(Q)과 비교하여, 신뢰도 손실( $L_{conf}(G, F)$ )을 계산한다. 이때, 학습부(400)는 추정 신뢰도 맵( $F(G(C), I^1)$ )의 각 픽셀값과 진리 신뢰도 맵(Q)의 대응하는 픽셀값을 비교하여, 추정 신뢰도 맵( $F(G(C), I^1)$ )의 각 픽셀값과 진리 신뢰도 맵의 대응하는 픽셀값이 동일하면 양성(positive)으로 판별하고, 서로 상이하면 음성(negative)로 판별하여 판별 결과에 따라 신뢰도 손실( $L_{conf}(G, F)$ )을 구분하여 계산할 수 있다.

[0063] 이에 신뢰도 손실( $L_{conf}(G, F)$ )은 수학식 2와 같이 계산될 수 있다.

수학식 2

$$\mathcal{L}_{\text{conf}}(G, F) = \mathbb{E}_{\hat{C}^t \sim p_{\text{pos}}(C^t), \hat{D}^t \sim p_{\text{pos}}(D^t), I^l \sim p_{\text{pos}}(I^l)} [\log F(\hat{C}^t, \hat{D}^t, I^l)] \\ + \mathbb{E}_{C \sim p_{\text{neg}}(C), I^l \sim p_{\text{neg}}(I^l)} [\log (1 - F(G(C), I^l))]$$

[0064]

[0065]

여기서  $\hat{C}^t$ 와  $\hat{D}^t$ 는 각각 t번째 반복 학습에서 양성으로 판별된 코스트 볼륨과 디스패리티 맵을 나타내고,  $\mathbb{E}_{\hat{C}^t \sim p_{\text{pos}}(C^t), \hat{D}^t \sim p_{\text{pos}}(D^t), I^l \sim p_{\text{pos}}(I^l)}$ 는 t번째 반복 학습에서 코스트 볼륨( $C^t$ )과 디스패리티 맵( $D^t$ ) 및 컬러 영상( $I^l$ ) 각각의 양성으로 판별된 코스트 볼륨( $\hat{C}^t$ )과 디스패리티 맵( $\hat{D}^t$ ) 및 컬러 영상( $I^l$ ) 픽셀의 크로스 엔트로피를 나타내는 에너지 함수이며,  $\mathbb{E}_{C \sim p_{\text{neg}}(C), I^l \sim p_{\text{neg}}(I^l)}$ 는 코스트 볼륨(C)과 컬러 영상( $I^l$ )에서 음성으로 판별된 코스트 볼륨(C)과 컬러 영상( $I^l$ ) 픽셀의 크로스 엔트로피를 나타내는 에너지 함수이다.

[0066]

수학식 2에서 음성으로 판별된 디스패리티 맵(D)에 대한 크로스 엔트로피를 계산하지 않는 것은 판별자인 신뢰도 판정부(300)가 이미 신뢰도가 없다고 판정한 결과이기 때문이다.

[0067]

수학식 2에서는 신뢰도 손실( $\mathcal{L}_{\text{conf}}(G, F)$ )을 양성과 음성으로 구분하여 계산하였으나, 여기서 양성으로 판정된 결과는 디스패리티 맵 획득부(220)가 정상적으로 디스패리티 맵( $G(C)$ )을 추정하고, 신뢰도 판정부(300)가 정상적으로 신뢰도가 있는 것으로 판정한 결과로 볼 수 있다. 양성 판별 결과는 신뢰도 손실( $\mathcal{L}_{\text{conf}}(G, F)$ )로 보기 어려우며, 따라서 본 실시예에서 학습부(400)는 양성과 음성으로 구분하여 계산된 신뢰도 손실( $\mathcal{L}_{\text{conf}}(G, F)$ ) 중 음성으로 계산된 신뢰도 손실만을 수학식 3과 같이 획득할 수 있다.

수학식 3

$$\mathcal{L}_{\text{conf}}(G, F) = \mathbb{E}_{C \sim p_{\text{neg}}(C), I^l \sim p_{\text{neg}}(I^l)} [\log (1 - F(G(C), I^l))]$$

[0068]

[0069]

학습부(400)는 수학식 1에 따라 디스패리티 손실( $\mathcal{L}_{\text{disp}}(G)$ )이 계산되고 수학식 3에 따라 신뢰도 손실( $\mathcal{L}_{\text{conf}}(G, F)$ )이 계산되면, 수학식 4와 같이 총 손실( $\mathcal{L}_{\text{total}}(G, F)$ )을 계산하여 역전과하여 인공 신경망으로 구성되는 디스패리티 맵 획득부(220)와 신뢰도 판정부(300)를 학습시킬 수 있다.

수학식 4

$$\mathcal{L}_{\text{total}}(G, F) = \mathcal{L}_{\text{disp}}(G) + \lambda \mathcal{L}_{\text{conf}}(G, F)$$

[0070]

[0071]

이때, 디스패리티 맵 획득부(220)와 신뢰도 판정부(300)는 생성적 적대 신경망 구조로 구성되어 학습되므로, 디스패리티 맵 획득부(220)와 신뢰도 판정부(300)의 네트워크 함수( $G, F$ )는 수학식 5로 최적화될 수 있다.

수학식 5

$$\{G^*, F^*\} = \underset{G}{\operatorname{argmin}} \mathcal{L}_{\text{disp}}(G) + \lambda \underset{F}{\operatorname{argmax}} \mathcal{L}_{\text{conf}}(G, F) \\ + \lambda \underset{G}{\operatorname{argmax}} \mathbb{E}_{C \sim p_{\text{neg}}(C), I^l \sim p_{\text{neg}}(I^l)} [\log F(G(C), I^l)]$$

[0072]

[0073]

상기에서 학습부(400)는 디스패리티 맵 획득부(220)와 신뢰도 판정부(300)를 학습시키기 위한 구성으로 학습이

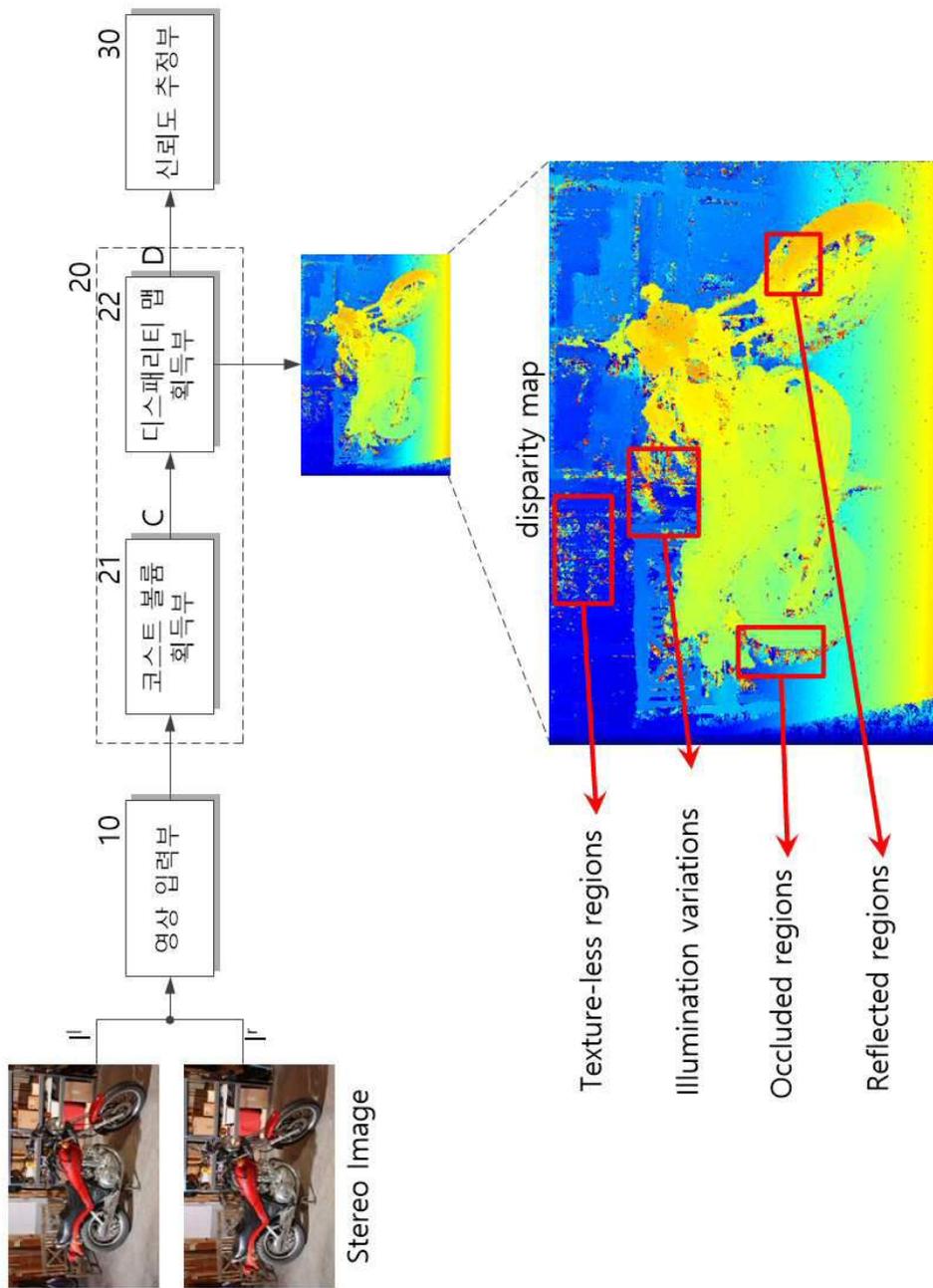
완료된 이후에는 생략될 수 있다. 경우에 따라서는 신뢰도 판정부(300) 또한 학습이 완료된 이후에는 제외되어 디스패리티 맵만을 획득하도록 구성될 수도 있다.

- [0074] 도 8은 본 실시예에 따른 스테레오 매칭 신뢰도 추정 장치의 성능을 비교한 결과를 나타낸다.
- [0075] 도 8에서 (a)는 스테레오 영상의 좌 영상( $I^l$ )을 나타내고, (b)는 기존의 인공 신경망을 이용하여 추정된 초기 디스패리티 맵이며, (c) 및 (d)는 일반 학습 방식으로 학습되어 획득된 디스패리티 맵 및 신뢰도 맵을 나타내며, (e) 및 (f)는 본 실시예에 따른 생성적 적대 방식으로 학습되어 획득된 디스패리티 맵 및 신뢰도 맵을 나타낸다.
- [0076] 도 8의 (c)와 (d)를 (e)와 (f)를 비교하면, 본 실시예의 생성적 적대 방식으로 학습된 스테레오 매칭 신뢰도 추정 장치가 더 정확한 디스패리티 맵과 신뢰도 맵을 획득할 수 있음을 알 수 있다.
- [0077] 도 9는 본 발명의 일 실시예에 따른 스테레오 매칭 신뢰도 추정 방법을 나타낸다.
- [0078] 도 4 내지 도 7을 참조하여, 도 9의 스테레오 매칭 신뢰도 추정 방법을 설명하면, 스테레오 매칭 신뢰도 추정 방법은 크게 학습 단계(S10)와 신뢰도 추정 단계(S20)로 구분될 수 있다.
- [0079] 학습 단계(S10)에서는 우선 학습 스테레오 영상이 입력된다(S11). 여기서 학습 스테레오 영상은 진리 디스패리티 맵( $D^*$ )이 함께 획득된 스테레오 영상을 의미한다.
- [0080] 학습 스테레오 영상이 입력되면, 학습 스테레오 영상의 2개의 영상( $I^l, I^r$ ) 중 하나를 기준 영상( $I^l$ )으로 하여, 기준 영상의 각 픽셀에 대한 나머지 영상의 픽셀들과의 매칭 코스트를 기지정된 방식으로 계산하여 코스트 볼륨(C)을 획득한다(S12).
- [0081] 그리고 인공 신경망을 이용하여 코스트 볼륨(C)로부터 현재까지 학습된 방식에 따라 추정 디스패리티 맵( $G(C)$ )을 추정한다(S13). 이때 추정 디스패리티 맵( $G(C)$ )은 코스트 볼륨(C)을 학습된 방식에 따라 인코딩 및 디코딩하여, 각 픽셀별 매칭 코스트의 확률 밀도를 추출하고, 추출된 확률 밀도를 정규화하여 정규 코스트 볼륨( $C_n$ )을 획득하며, 획득된 정규 코스트 볼륨( $C_n$ )의 각 픽셀별로 정규화된 다수의 매칭 코스트 중 최대가 되는 매칭 코스트를 나타내는 디스패리티(d)를 추출함으로써 추정 디스패리티 맵( $G(C)$ )을 추정할 수 있다.
- [0082] 한편, 정규 코스트 볼륨( $C_n$ )에 대해 학습되는 방식에 따라 기설정된 K개의 코스트를 추출 및 폴링하여 정제 코스트 볼륨( $C_r$ )을 획득한다(S14).
- [0083] 추정 디스패리티 맵( $G(C)$ )과 정제 코스트 볼륨( $C_r$ )이 획득되면, 인공 신경망을 이용하여 디스패리티 맵( $G(C)$ )과 정제 코스트 볼륨( $C_r$ ) 및 기준 영상( $I^l$ ) 각각에 대해 학습되는 방식에 따라 특징을 추출하여 코스트 특징맵과 디스패리티 특징맵 및 컬러 특징맵을 획득한다(S15). 그리고 획득된 코스트 특징맵과 디스패리티 특징맵 및 컬러 특징맵을 기지정된 방식으로 결합하여 결합 특징맵을 획득한다(S16). 결합 특징맵이 획득되면, 학습되는 방식에 따라 획득된 결합 특징맵의 특징을 추출하여 추정 신뢰도 맵( $F(G(C))$ )을 획득한다(S17).
- [0084] 추정 신뢰도 맵( $F(G(C))$ )이 획득되면, 추정 디스패리티 맵( $G(C)$ )과 진리 디스패리티 맵( $D^*$ ) 및 추정 신뢰도 맵( $F(G(C))$ )을 기반으로 추정 신뢰도 맵( $F(G(C))$ )의 각 픽셀별 추정 신뢰도가 정상적으로 추정 디스패리티 맵( $G(C)$ )의 신뢰도를 정상적으로 판정한 결과인지 여부에 따라 양성 또는 음성을 판별한다(S18).
- [0085] 이후, 추정 디스패리티 맵( $G(C)$ )의 생성 과정에서 발생한 디스패리티 손실( $L_{disp}(G)$ )을 수학적 식 1에 따라 계산하여 획득하고, 추정 디스패리티 맵( $G(C)$ )의 각 픽셀별 신뢰도를 적대적으로 판정하는 과정에서 발생한 신뢰도 손실( $L_{conf}(G,F)$ ) 중 음성으로 판별된 손실을 수학적 식 3에 따라 계산하고, 계산된 디스패리티 손실( $L_{disp}(G)$ )과 손실( $L_{conf}(G,F)$ )을 결합하여 수학적 식 4로 계산되는 총 손실( $L_{total}(G,F)$ )을 역전파하여 인공 신경망을 학습시킨다(S19). 여기서는 설명의 편의를 위하여 생략하였으나, 학습 단계(S10)는 인공 신경망의 네트워크 함수( $G,F$ )가 수학적 식 5에 대응하여 최적화되도록 반복 수행될 수 있다.
- [0086] 한편, 학습이 완료된 이후, 신뢰도 추정 단계(S20)에서도 우선 스테레오 영상이 입력된다(S21). 여기서 스테레오 영상은 학습 스테레오 영상과 달리 진리 디스패리티 맵( $D^*$ )이 획득되지 않은 스테레오 영상을 나타낸다.

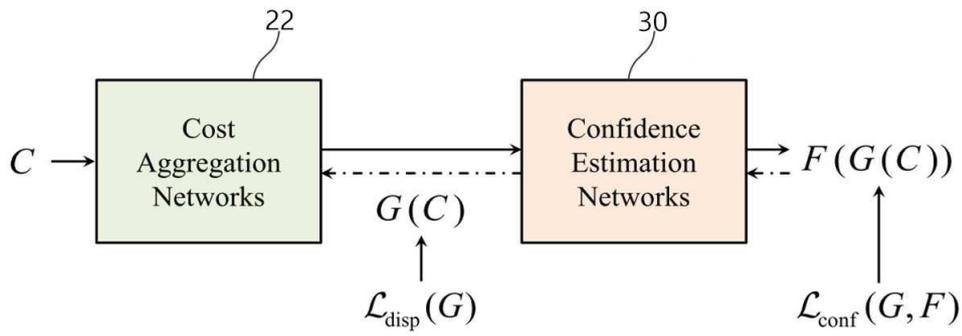


도면

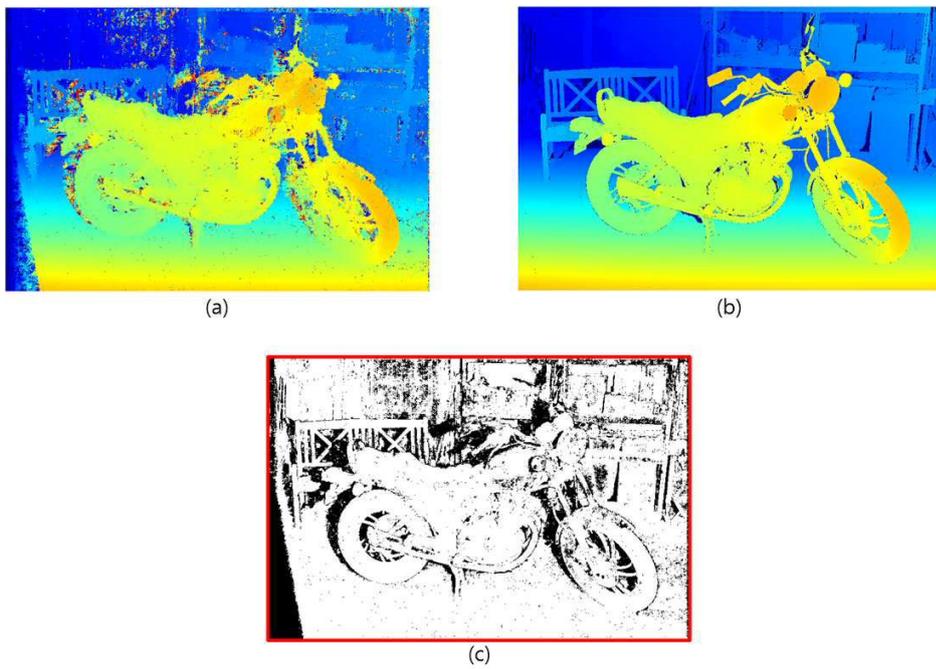
도면1



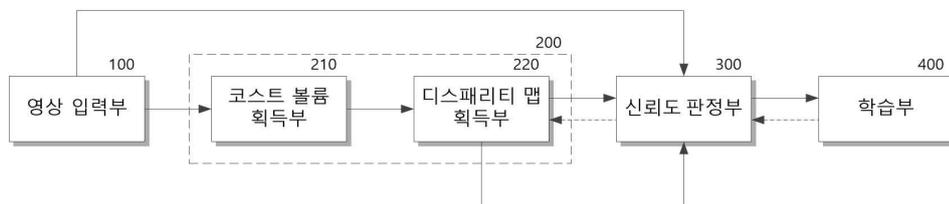
도면2



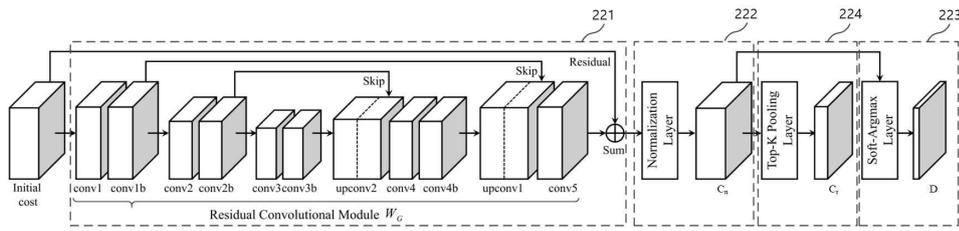
도면3



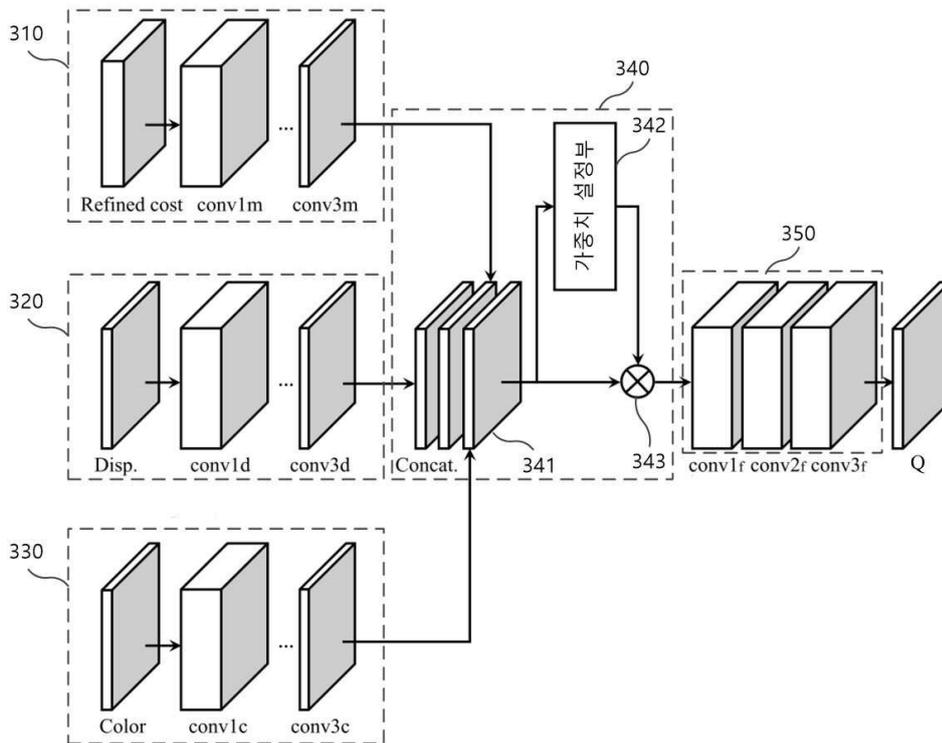
도면4



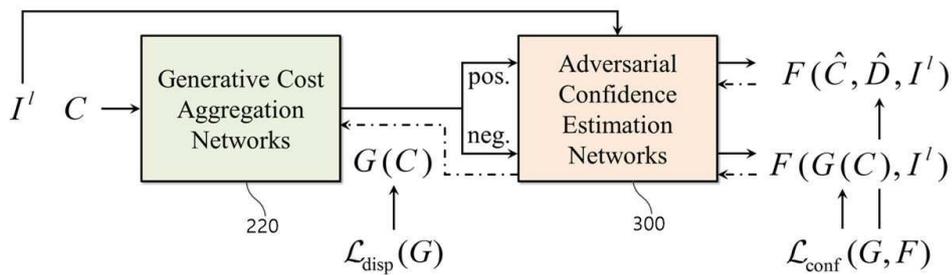
도면5



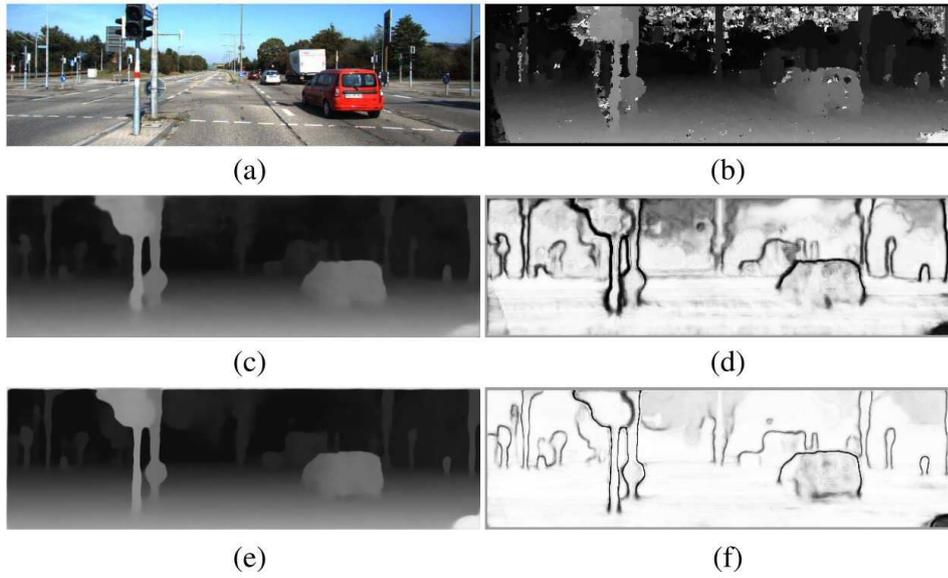
도면6



도면7



도면8



도면9

