(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2021년06월08일
(11) 등록번호 10-2262832
(24) 등록일자 2021년06월03일

(51) 국제특허분류(Int. Cl.)
G06T 7/50 (2017.01) *G06T 3/00* (2019.01)
(52) CPC특허분류
G06T 7/50 (2017.01)
G06N 3/08 (2013.01)
(21) 출원번호 10-2019-0156863
(22) 출원일자 2019년11월29일
심사청구일자 2019년11월29일
(56) 선행기술조사문헌
Michele Mancini 등. Toward Domain Independence for Learning-Based Monocular Depth Estimation, IEEE Robotics and Automation Letters.(2017.07.)*
(뒷면에 계속)

(73) 특허권자
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
(72) 발명자
함범섭
서울특별시 강남구 압구정로61길 37, 72동 506호 (압구정동, 한양아파트)
엄찬호
서울특별시 마포구 백범로 230, 102동 2203호(신공덕동, 브라운스톤 공덕 아파트)
박현종
서울특별시 서대문구 연희로 82, A동 612호(연희동)
(74) 대리인
민영준

전체 청구항 수 : 총 13 항 심사관 : 경연정

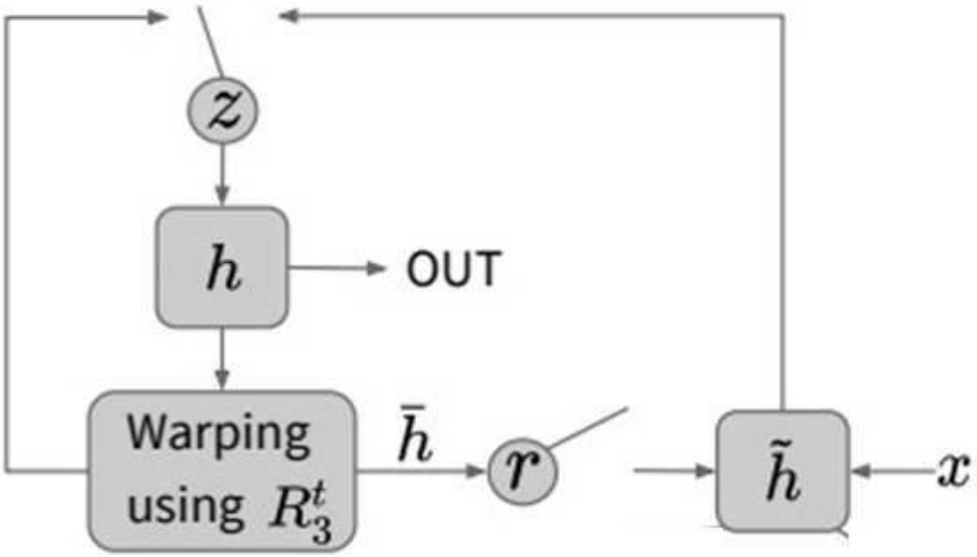
(54) 발명의 명칭 **단안 비디오 영상의 깊이 추정 방법 및 장치**

(57) 요 약

단안 비디오 영상의 깊이 추정 방법 및 장치가 개시된다. 개시된 장치는, 현재 프레임 영상에 대한 신경망 연산을 통해 공간 특징맵을 생성하는 공간 특징 엔코더 네트워크 모듈; 현재 프레임 영상과 이전 프레임 영상의 광학 플로우 영상에 대한 신경망 연산을 통해 시간 특징맵을 생성하는 시간 특징 엔코더 네트워크 모듈; 상기 공간 특

(뒷면에 계속)

대 표 도 - 도4



징맵 및 상기 시간 특징맵을 이용하여 상기 현재 프레임 영상에 대한 깊이 특징맵을 신경망 연산을 통해 생성하는 플로우 가이드 메모리 모듈; 및 상기 깊이 특징맵에 대한 신경망 연산을 통해 깊이 맵을 생성하는 디코더 네트워크 모듈을 포함하되, 상기 플로우 가이드 메모리 모듈은 RNN을 사용하며, 상기 RNN에 사용되는 이전 상태 특징맵을 상기 시간 특징맵에 기초한 워핑을 통해 보정하고, 상기 이전 상태 특징맵 대신 상기 보정된 이전 상태 특징맵을 이용하여 신경망 연산을 수행한다. 개시된 장치 및 방법에 의하면, 단안 비디오 영상에서 프레임간 상관 관계를 고려하여 정확하게 깊이를 추정할 수 있는 장점이 있다.

(52) CPC특허분류
G06T 3/0093 (2013.01)
G06T 2207/10028 (2013.01)
G06T 2207/20084 (2013.01)

(56) 선행기술조사문헌
이승수 등. 깊이맵 생성 알고리즘의 합성곱 신경망 구현, 방송공학회논문지.(2018.01)*
Yang Feng 등. Spatio-temporal Video Re-localization by Warp LSTM, arXiv:1905.03922v1.(2019.05.10.)*
US20190279383 A1
KR1020150079576 A
*는 심사관에 의하여 인용된 문헌

이 발명을 지원한 국가연구개발사업
과제고유번호 20160001970041001
부처명 과학기술정보통신부
과제관리(전문)기관명 정보통신기획평가원(한국연구재단부설)
연구사업명 정보통신방송연구개발사업
연구과제명 스마트카 다중 센서와 딥러닝을 이용한 초정밀 내츄럴 3D 뷰 생성 기술 개발 (창조씨앗형 2단계)(3/5)
기 여 율 1/1
과제수행기관명 연세대학교 산학협력단
연구기간 2019.01.01 ~ 2019.12.31
공지예외적용 : 있음

명 세 서

청구범위

청구항 1

현재 프레임 영상에 대한 신경망 연산을 통해 공간 특징맵을 생성하는 공간 특징 엔코더 네트워크 모듈;

현재 프레임 영상과 이전 프레임 영상의 광학 플로우 영상에 대한 신경망 연산을 통해 시간 특징맵을 생성하는 시간 특징 엔코더 네트워크 모듈;

상기 공간 특징맵 및 상기 시간 특징맵을 이용하여 상기 현재 프레임 영상에 대한 깊이 특징맵을 신경망 연산을 통해 생성하는 플로우 가이드 메모리 모듈; 및

상기 깊이 특징맵에 대한 신경망 연산을 통해 깊이 맵을 생성하는 디코더 네트워크 모듈을 포함하되,

상기 플로우 가이드 메모리 모듈은 RNN을 사용하며, 상기 RNN에 사용되는 이전 상태 특징맵을 상기 시간 특징맵에 기초한 워핑을 통해 보정하고, 상기 이전 상태 특징맵 대신 상기 보정된 이전 상태 특징맵을 이용하여 신경망 연산을 수행하며,

상기 현재 프레임 영상, 상기 광학 플로우 영상 및 이전 프레임 영상에 대한 신경망 연산을 통해 교정된 시간 특징맵을 생성하는 광학 플로우 교정 네트워크 모듈을 포함하는 것을 특징으로 하는 단안 비디오 영상의 깊이 추정 장치.

청구항 2

삭제

청구항 3

제1항에 있어서,

상기 플로우 가이드 메모리 모듈은 상기 시간 특징맵 대신 상기 교정된 시간 특징맵에 기초한 워핑을 통해 상기 이전 상태 특징맵을 보정하는 것을 특징으로 하는 단안 비디오 영상의 깊이 추정 장치.

청구항 4

제3항에 있어서,

상기 보정되는 이전 상태 특징맵은 마스크 특징맵에 의해 그 값이 조절되며, 상기 마스크 특징맵은 상기 시간 특징맵 또는 상기 교정된 시간 특징맵의 신뢰도를 반영한 특징맵인 것을 특징으로 하는 단안 비디오 영상의 깊이 추정 장치.

청구항 5

제4항에 있어서,

상기 신뢰도는 상기 시간 특징맵 또는 상기 교정된 시간 특징맵에 기초하여 이전 프레임 영상을 워핑한 영상과 현재 프레임 영상간의 차에 기초하여 연산되는 것을 특징으로 하는 단안 비디오 영상의 깊이 추정 장치.

청구항 6

제1항에 있어서,

상기 플로우 가이드 메모리 모듈의 RNN은 다음의 수학식과 같이 현재 상태 특징맵(h^t), 보정된 이전 상태 특징맵($\bar{h}^t$), 리셋 게이트(r^t) 및 업데이트 게이트(z^t) 및 후보 상태 특징맵($\tilde{h}^t$)을 연산하는 것을 특징으로 하는 단안 비디오 영상의 깊이 추정 장치.

$$\bar{h}^t = M^t \odot \mathcal{W}(h^{t-1}; R_3^t)$$
$$z^t = \sigma(W_{xz} * x^t + W_{\bar{h}z} * \bar{h}^t + b_z)$$
$$r^t = \sigma(W_{xr} * x^t + W_{\bar{h}r} * \bar{h}^t + b_r)$$
$$\tilde{h}^t = \tanh(W_{x\tilde{h}} * x^t + r^t \odot (W_{\bar{h}\tilde{h}} * \bar{h}^t) + b_{\tilde{h}})$$
$$h^t = (1 - z^t) \odot \bar{h}^t + z^t \odot \tilde{h}^t,$$

위 수학식에서, σ는 시그모이드 함수를 의미하고, $\odot$는 엘리먼트-와이즈(element-wise) 곱셉을 의미하며, *는 콘볼루션을 의미하며, x^t는 입력되는 특징맵으로서, 공간 특징맵과 시간 특징맵을 결합한 특징맵이고, W는 미리 설정되는 가중치이며, b는 미리 설정되는 바이어스 값이고, $\mathcal{W}(h^{t-1}; R_3^t)$ 는 시간 특징맵 또는 교정된 시간 특징맵을 이용하여 이전 상태 특징맵을 워핑한 특징맵이고, M^t는 마스크 특징맵임.

청구항 7

제4항에 있어서,

상기 마스크 특징맵은 다음의 수학식과 같이 설정되는 것을 특징으로 하는 단안 비디오 영상의 깊이 추정 장치.

$$M^t(p) = \exp(-\epsilon \| I_3^t(p) - \bar{I}_3^t(p) \|_1)$$

위 수학식에서, p는 픽셀을 의미하고, $I^t_3(p)$는 현재 프레임 영상이며, $\bar{I}^t_3(p)$는 워핑된 이전 프레임 영상이고, ε 은 임의로 설정되는 상수임.

청구항 8

제1항에 있어서,

상기 공간 특징 엔코더 네트워크 모듈 및 상기 시간 특징 엔코더 네트워크 모듈은 CNN을 이용하여 각각 공간 특징맵 및 시간 특징맵을 생성하는 것을 특징으로 하는 단안 비디오 영상의 깊이 추정 장치.

청구항 9

현재 프레임 영상에 대한 신경망 연산을 통해 공간 특징맵을 생성하는 단계(a);

현재 프레임 영상과 이전 프레임 영상의 광학 플로우 영상에 대한 신경망 연산을 통해 시간 특징맵을 생성하는 단계(b);

상기 공간 특징맵 및 상기 시간 특징맵을 이용하여 상기 현재 프레임 영상에 대한 깊이 특징맵을 신경망 연산을 통해 생성하는 단계(c); 및

상기 깊이 특징맵에 대한 신경망 연산을 통해 깊이 맵을 생성하는 단계(d)를 포함하되,

상기 단계(c)는 RNN을 사용하며, 상기 RNN에 사용되는 이전 상태 특징맵을 상기 시간 특징맵에 기초한 워핑을 통해 보정하고, 상기 이전 상태 특징맵 대신 상기 보정된 이전 상태 특징맵을 이용하여 신경망 연산을 수행하며,

상기 현재 프레임 영상, 상기 광학 플로우 영상 및 이전 프레임 영상에 대한 신경망 연산을 통해 교정된 시간 특징맵을 생성하는 단계를 더 포함하는 것을 특징으로 하는 단안 비디오 영상의 깊이 추정 방법.

청구항 10

삭제

청구항 11

제9항에 있어서,

상기 단계(c)는 상기 시간 특징맵 대신 상기 교정된 시간 특징맵에 기초한 워핑을 통해 상기 이전 상태 특징맵을 보정하는 것을 특징으로 하는 단안 비디오 영상의 깊이 추정 방법.

청구항 12

제11항에 있어서,

상기 보정되는 이전 상태 특징맵은 마스크 특징맵에 의해 그 값이 조절되며, 상기 마스크 특징맵은 상기 시간 특징맵 또는 상기 교정된 시간 특징맵의 신뢰도를 반영한 특징맵인 것을 특징으로 하는 단안 비디오 영상의 깊이 추정 방법.

청구항 13

제12항에 있어서,

상기 신뢰도는 상기 시간 특징맵 또는 상기 교정된 시간 특징맵에 기초하여 이전 프레임 영상을 워핑한 영상과 현재 프레임 영상간의 차에 기초하여 연산되는 것을 특징으로 하는 단안 비디오 영상의 깊이 추정 방법.

청구항 14

제9항에 있어서,

상기 RNN은 다음의 수학식과 같이 현재 상태 특징맵(h^t), 보정된 이전 상태 특징맵($\bar{h}^t$), 리셋 게이트(r^t) 및 업데이트 게이트(z^t) 및 후보 상태 특징맵($\tilde{h}^t$)을 연산하는 것을 특징으로 하는 단안 비디오 영상의 깊이 추정 방법.

$$\bar{h}^t = M^t \odot \mathcal{W}(h^{t-1}; R_3^t)$$
$$z^t = \sigma(W_{xz} * x^t + W_{\bar{h}z} * \bar{h}^t + b_z)$$
$$r^t = \sigma(W_{xr} * x^t + W_{\bar{h}r} * \bar{h}^t + b_r)$$
$$\tilde{h}^t = \tanh(W_{x\tilde{h}} * x^t + r^t \odot (W_{\bar{h}\tilde{h}} * \bar{h}^t) + b_{\tilde{h}})$$

$$h^t = (1 - z^t) \odot \bar{h}^t + z^t \odot \tilde{h}^t,$$

위 수학식에서, σ는 시그모이드 함수를 의미하고, $\odot$는 엘리먼트-와이즈(element-wise) 곱셉을 의미하며, *는 콘볼루션을 의미하며, x^t는 입력되는 특징맵으로서, 공간 특징맵과 시간 특징맵을 결합한 특징맵이고, W는 미리 설정되는 가중치이며, b는 미리 설정되는 바이어스 값이고, $\mathcal{W}(h^{t-1}; R_3^t)$ 는 시간 특징맵 또는 교정된 시간 특징맵을 이용하여 이전 상태 특징맵을 워핑한 특징맵이고, M^t는 마스크 특징맵임.

청구항 15

제12항에 있어서,

상기 마스크 특징맵은 다음의 수학식과 같이 설정되는 것을 특징으로 하는 단안 비디오 영상의 깊이 추정 방법.

$$M^t(p) = \exp(-\epsilon \| I_3^t(p) - \bar{I}_3^t(p) \|_1)$$

위 수학식에서, p는 픽셀을 의미하고, $I^t_3(p)$는 현재 프레임 영상이며, $\bar{I}^t_3(p)$는 워핑된 이전 프레임 영상이고, ε 은 임의로 설정되는 상수임.

발명의 설명

기 술 분 야

[0001] 본 발명은 깊이 추정 장치 및 방법에 관한 것으로서, 더욱 상세하게는 단안 비디오 영상의 깊이 추정 장치 및 방법에 관한 것이다.

배 경 기 술

[0003] 깊이 추정은 자율 주행 및 운전자 보조 시스템에서 필수적으로 필요한 기술이다. 자율 주행 시 지형 구조의 판단 및 주변 차량과 장애물의 정확한 위치 판단을 위해 실시간 깊이 추정이 필요하다.

[0004] 일반적인 깊이 추정은 스테레오 매칭을 통해 이루어진다. 스테레오 매칭은 두 개의 카메라를 이용하여 획득되는 좌영상 및 우영상을 이용하여 깊이를 추정하는 방법이다. 스테레오 매칭에서는 좌영상과 우영상의 대응 픽셀간의 변이를 연산하여 깊이를 추정하게 된다.

[0005] 그러나, 스테레오 매칭은 항상 두 개의 카메라를 이용하여 영상을 획득하여야 하며 두 개의 카메라가 정확히 정렬된 상태에서 영상을 획득하여야 정확한 깊이 추정이 가능한 문제가 있어 현실적으로 사용하기 어려운 측면이 있다.

[0006] 단안 카메라를 이용하여 깊이를 추정하는 방법에 대해서도 다양한 방법들이 제안되었다. 객체의 상대적 크기, 텍스터 변화 정도, 가려진 영역 등은 단안 영상에서도 깊이를 추정할 수 있는 정보를 제공하고 이러한 정보를 이용하게 될 경우 스테레오 영상이 주어지지 않더라도 깊이 추정이 가능하다.

[0007] 한편, 근래에 들어 딥러닝에 대한 연구가 발전하면서 신경망 연산을 통해 단안 영상의 깊이를 추정하는 다양한 방법들이 제안되고 있다.

[0008] 그러나, 기존의 신경망을 이용한 단안 영상의 깊이 추정 방법은 프레임 단위로 독립적으로 깊이를 추정하였다. 연속되는 프레임은 서로 높은 상관 관계가 있음에도 불구하고 프레임간 상관 관계는 단안 영상의 깊이 추정에 잘 반영되지 않았으며, 이는 비디오 영상에서 부정확한 깊이 추정의 주요한 원인 중 하나가 되었다.

발명의 내용

해결하려는 과제

[0010] 본 발명은 단안 비디오 영상에서 프레임간 상관 관계를 고려하여 정확하게 깊이를 추정할 수 있는 깊이 추정 장치 및 방법을 제안한다.

과제의 해결 수단

[0012] 상기와 같은 목적을 달성하기 위해, 본 발명의 일 측면에 따르면, 현재 프레임 영상에 대한 신경망 연산을 통해 공간 특징맵을 생성하는 공간 특징 엔코더 네트워크 모듈; 현재 프레임 영상과 이전 프레임 영상의 광학 플로우 영상에 대한 신경망 연산을 통해 시간 특징맵을 생성하는 시간 특징 엔코더 네트워크 모듈; 상기 공간 특징맵 및 상기 시간 특징맵을 이용하여 상기 현재 프레임 영상에 대한 깊이 특징맵을 신경망 연산을 통해 생성하는 플로우 가이드 메모리 모듈; 및 상기 깊이 특징맵에 대한 신경망 연산을 통해 깊이 맵을 생성하는 디코더 네트워크 모듈을 포함하되, 상기 플로우 가이드 메모리 모듈은 RNN을 사용하며, 상기 RNN에 사용되는 이전 상태 특징맵을 상기 시간 특징맵에 기초한 워핑을 통해 보정하고, 상기 이전 상태 특징맵 대신 상기 보정된 이전 상태 특징맵을 이용하여 신경망 연산을 수행하는 단안 비디오 영상의 깊이 추정 장치가 제공된다.

[0013] 상기 현재 프레임 영상, 상기 광학 플로우 영상 및 이전 프레임 영상에 대한 신경망 연산을 통해 교정된 시간 특징맵을 생성하는 광학 플로우 교정 네트워크 모듈을 포함한다.

[0014] 상기 플로우 가이드 메모리 모듈은 상기 시간 특징맵 대신 상기 교정된 시간 특징맵에 기초한 워핑을 통해 상기 이전 상태 특징맵을 보정한다.

[0015] 상기 보정되는 이전 상태 특징맵은 마스크 특징맵에 의해 그 값이 조절되며, 상기 마스크 특징맵은 상기 시간 특징맵 또는 상기 교정된 시간 특징맵의 신뢰도를 반영한 특징맵이다. ,

[0016] 상기 신뢰도는 상기 시간 특징맵 또는 교정된 시간 특징맵에 기초하여 이전 프레임 영상을 워핑한 영상과 현재 프레임 영상간의 차에 기초하여 연산된다.

[0017] 상기 플로우 가이드 메모리 모듈의 RNN은 다음의 수학식과 같이 현재 상태 특징맵(h^t), 보정된 이전 상태 특징맵($\bar{h}^t$), 리셋 게이트(r^t) 및 업데이트 게이트(z^t) 및 후보 상태 특징맵($\tilde{h}^t$)을 연산한다

$$\bar{h}^t = M^t \odot \mathcal{W}(h^{t-1}; R_3^t)$$
$$z^t = \sigma(W_{xz} * x^t + W_{\bar{h}z} * \bar{h}^t + b_z)$$
$$r^t = \sigma(W_{xr} * x^t + W_{\bar{h}r} * \bar{h}^t + b_r)$$
$$\tilde{h}^t = \tanh(W_{x\tilde{h}} * x^t + r^t \odot (W_{\bar{h}\tilde{h}} * \bar{h}^t) + b_{\tilde{h}})$$

[0018]
$$h^t = (1 - z^t) \odot \bar{h}^t + z^t \odot \tilde{h}^t,$$

[0019] 위 수학식에서, σ는 시그모이드 함수를 의미하고, $\odot$는 엘리먼트-와이즈(element-wise) 곱셉을 의미하며, *는 콘볼루션을 의미하며, x^t는 입력되는 특징맵으로서, 공간 특징맵과 시간 특징맵을 결합한 특징맵이고, W는 미리 설정되는 가중치이며, b는 미리 설정되는 바이어스 값이고, $\mathcal{W}(h^{t-1}; R_3^t)$ 는 시간 특징맵 또는 교정된 시간 특징맵을 이용하여 이전 상태 특징맵을 워핑한 특징맵이고, M^t는 마스크 특징맵임.

[0020] 상기 마스크 특징맵은 다음의 수학식과 같이 설정된다.

[0021]
$$M^t(p) = \exp(-\epsilon \| I_3^t(p) - \bar{I}_3^t(p) \|_1)$$

[0022] 위 수학식에서, p는 픽셀을 의미하고, $I^t_3(p)$는 현재 프레임 영상이며, $\bar{I}^t_3(p)$는 워핑된 이전 프레임 영상이고, ε 은 임의로 설정되는 상수임.

[0023] 상기 공간 특징 엔코더 네트워크 모듈 및 상기 시간 특징 엔코더 네트워크 모듈은 CNN을 이용하여 각각 공간 특징맵 및 시간 특징맵을 생성한다.

[0024] 본 발명의 다른 측면에 따르면, 현재 프레임 영상에 대한 신경망 연산을 통해 공간 특징맵을 생성하는 단계(a); 현재 프레임 영상과 이전 프레임 영상의 광학 플로우 영상에 대한 신경망 연산을 통해 시간 특징맵을 생성하는

단계(b); 상기 공간 특징맵 및 상기 시간 특징맵을 이용하여 상기 현재 프레임 영상에 대한 깊이 특징맵을 신경망 연산을 통해 생성하는 단계(c); 및 상기 깊이 특징맵에 대한 신경망 연산을 통해 깊이 맵을 생성하는 단계(d)를 포함하되, 상기 단계(c)는 RNN을 사용하며, 상기 RNN에 사용되는 이전 상태 특징맵을 상기 시간 특징맵에 기초한 워핑을 통해 보정하고, 상기 이전 상태 특징맵 대신 상기 보정된 이전 상태 특징맵을 이용하여 신경망 연산을 수행하는 단안 비디오 영상의 깊이 추정 방법이 제공된다.

발명의 효과

[0026] 본 발명에 의하면, 단안 비디오 영상에서 프레임간 상관 관계를 고려하여 정확하게 깊이를 추정할 수 있는 장점이 있다.

도면의 간단한 설명

[0028] 도 1은 본 발명의 제1 실시에에 따른 단안 비디오 영상의 깊이 추정 장치를 구성하는 뉴럴 네트워크 구조를 도시한 도면.

도 2는 본 발명의 제2 실시예에 따른 단안 비디오 영상의 깊이 추정을 위한 뉴럴 네트워크 구조를 나타낸 도면.

도 3은 본 발명의 일 실시예에 따른 광학 플로우 교정 네트워크 모듈의 동작 구조를 나타낸 도면.

도 4는 본 발명의 일 실시예에 따른 플로우 가이드 메모리 모듈의 동작 구조를 나타낸 도면.

도 5는 본 발명의 일 실시예에 따른 플로우 가이드 메모리 모듈에서의 워핑을 개념적으로 나타낸 도면.

도 6은 본 발명의 제2 실시예에 따른 단안 비디오 영상의 깊이 추정 방법의 전체적인 흐름을 도시한 순서도.

발명을 실시하기 위한 구체적인 내용

[0029] 이하에서는 첨부한 도면을 참조하여 본 발명을 설명하기로 한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며, 따라서 여기에서 설명하는 실시예로 한정되는 것은 아니다.

[0030] 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

[0031] 명세서 전체에서, 어떤 부분이 다른 부분과 "연결"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우뿐 아니라, 그 중간에 다른 부재를 사이에 두고 "간접적으로 연결"되어 있는 경우도 포함한다.

[0032] 또한 어떤 부분이 어떤 구성 요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성 요소를 제외하는 것이 아니라 다른 구성 요소를 더 구비할 수 있다는 것을 의미한다.

[0033] 이하 첨부된 도면을 참고하여 본 발명의 실시예를 상세히 설명하기로 한다.

[0034] 본 발명의 깊이 추정 장치 및 방법은 단안 비디오 영상을 이용하여 단안 비디오 영상의 각 픽셀에 대한 깊이를 추정한다. 단안 비디오 영상은 스테레오 영상과는 달리 양안 시차 정보를 알 수 없기에 정확한 깊이 추정은 어려우나 항상 스테레오 영상을 획득할 수 없고 또한 정확히 정렬된 스테레오 영상을 획득하는 것 역시 용이하지 않은 문제이기에 단안 비디오 영상에 대한 깊이 추정은 계속적으로 요구되고 있으며, 특히 자율 주행과 같이 실시간으로 깊이 정보를 획득하여야 하는 분야에서 요구되고 있다.

[0035] 비디오 영상은 다수의 프레임으로 이루어져 있으며 인접하는 프레임(예를 들어, t-1 프레임, t 프레임, t+1 프레임)은 상호 상관성이 높다. 그러나, 비디오 영상에서의 깊이 추정은 각 프레임별로 이루어진다. 깊이 추정이 각 프레임별로 독립적으로 이루어지기 때문에 이전 프레임(t-1 프레임)에서 추정된 깊이는 현재 프레임(t 프레임)에서의 깊이 추정에 영향을 주지 않는다. 그런데, 인접하는 프레임이 서로 상관 관계에 있기 때문에 이전 프레임과 현재 프레임의 깊이 역시 상관 관계가 있으나 이러한 상관 관계는 기존의 단안 비디오 영상의 깊이 추정에는 고려되지 않았다. 이러한 문제는 기존의 단안 비디오 영상의 깊이 추정으로 인해 생성되는 깊이 맵에 플리커 현상을 발생시키는 원인이 되기도 하였다.

[0036] 본 발명에서 제안하는 단안 비디오 영상의 깊이 추정 장치 및 방법은 이러한 프레임간 깊이의 상관 관계가 반영될 수 있는 구성을 제안한다. 다만, 주목하여야 할 점은 비디오는 동영상이기 때문에 프레임 사이에 움직이는 객체가 존재할 수 있고 또한 카메라의 움직임으로 인해 뷰 포인트(View Point)가 변경될 수도 있다는 점이다. 이러한 객체 또는 카메라의 움직임을 반영하지 않고 단지 깊이의 상관 관계만을 고려하는 것은 오히려 부정확한

깊이 추정의 원인이 될 수도 있다. 본 발명은 이와 같은 객체 또는 카메라의 움직임을 함께 반영하여 프레임간 상관 관계가 반영된 깊이 추정 방법을 제안한다.

[0037] 도 1은 본 발명의 제1 실시예에 따른 단안 비디오 영상의 깊이 추정 장치를 구성하는 뉴럴 네트워크 구조를 도시한 도면이다.

[0038] 도 1을 참조하면, 본 발명의 제1 실시예에 따른 단안 비디오 영상의 깊이 추정 장치는 공간 특징 엔코더 네트워크 모듈(100), 시간 특징 엔코더 네트워크 모듈(110), 플로우 가이드 메모리 모듈(120) 및 디코더 네트워크 모듈(130)을 포함한다.

[0039] 본 발명의 단안 비디오 영상 깊이 추정 장치로는 두 개의 영상이 입력된다. 하나는 t 프레임의 입력 영상(I^t)이며 다른 하나는 t 프레임에서의 광학 플로우(Optical Flow) 영상(O^t)이다. 여기서, t 프레임에서의 광학 플로우 영상은 이전 프레임(t-1 프레임) 영상과 현재 프레임(t 프레임) 영상의 광학 플로우 연산 결과를 반영한 영상을 의미한다. 입력되는 광학 플로우 영상의 광학 플로우 값은 다양한 방법으로 연산될 수 있으며, 어떠한 방법으로 광학 플로우 영상을 생성하더라도 본 발명의 본질에는 영향을 미치지 아니한다. 널리 알려진 광학 플로우 연산 알고리즘이 이용될 수도 있을 것이며, 광학 플로우 연산을 위해 별도의 신경망이 이용될 수도 있다.

[0040] 공간 특징 엔코더 네트워크 모듈(100)은 신경망 연산을 통해 현재 프레임 영상에 대한 공간 특징맵을 생성한다. 공간 특징 엔코더 네트워크 모듈(100)은 공간 특징맵 생성을 위해 알려진 다양한 뉴럴 네트워크를 사용할 수 있다. 일례로, 공간 특징 엔코더 네트워크 모듈(100)은 콘볼루션 커널을 현재 프레임에 적용하면서 특징맵을 생성하는 CNN(Convolutional Neural Network) 네트워크를 포함할 수 있을 것이나 이에 한정되는 것은 아니다.

[0041] 공간 특징 엔코더 네트워크 모듈(100)은 입력 영상의 차원을 줄여나가면서 최종적인 특징맵을 생성할 수 있을 것이다. 일례로, 입력 영상에 비해 1/4 사이즈를 가지는 공간 특징맵이 공간 특징 엔코더를 통해 출력될 수 있을 것이다. 공간 특징 엔코더 네트워크 모듈(100)의 신경망 가중치는 학습을 통해 설정되며, 학습 방법은 추후 설명하기로 한다.

[0042] 시간 특징 엔코더 네트워크 모듈(110)은 신경망 연산을 통해 광학 플로우 영상에 대한 시간 특징맵을 생성한다. 시간 특징 네트워크 모듈(110)에서 출력되는 특징맵을 시간 특징맵으로 정의한 것은 광학 플로우의 정의가 시간에 따른 각 픽셀의 움직임 정도이기 때문이며, 시간 특징맵이 시간 정보 자체를 가리키는 것은 아니다.

[0043] 시간 특징 엔코더 네트워크 모듈(110)도 시간 특징맵 생성을 위해 알려진 다양한 뉴럴 네트워크를 사용할 수 있으며, 콘볼루션 커널을 광학 플로우 영상에 적용하면서 특징맵을 생성하는 CNN(Convolutional Neural Network) 네트워크를 사용할 수 있을 것이다.

[0044] 시간 특징 엔코더 네트워크 모듈(100) 역시 광학 플로우 영상의 차원을 줄여나가면서 최종적인 특징맵을 생성할 수 있을 것이며, 일례로, 입력된 광학 플로우 영상에 비해 1/4 사이즈를 가지는 시간 특징맵이 시간 특징 엔코더 네트워크 모듈(110)을 통해 출력될 수 있을 것이다. 시간 특징 엔코더 네트워크 모듈(110)의 신경망 가중치 역시 학습을 통해 설정된다.

[0045] 공간 특징맵 및 시간 특징맵은 상호 보완적인 특징 정보이다. 공간 특징맵은 영상의 존재하는 객체들의 형상 및 배경의 레이아웃들에 대한 특징을 포함하게 된다. 또한, 시간 특징맵은 프레임의 변화에 따른 각 픽셀들의 개별적인 움직임 궤적 정보를 포함하게 된다.

[0046] 본 발명의 바람직한 실시예에 따르면, 공간 특징 엔코더 네트워크 모듈(100) 및 시간 특징 엔코더 네트워크 모듈(110)에 적용되는 콘볼루션 연산 시 수용 영역(Receptive Field)을 확장시키기 위한 팽창된 콘볼루션(Dilated Convolution)을 사용하는 것이 바람직하다. 팽창된 콘볼루션(Dilated Convolution)을 사용하게 될 경우 공간 정보 해상도의 손실 및 장면 디테일의 손실을 최소화할 수 있어 본 발명과 같은 깊이 추정을 위한 신경망에 보다 효과적일 수 있다.

[0047] 팽창된 콘볼루션 연산 시 팽창 비율 및 이에 의존적인 수용 영역의 사이즈는 적절히 조절될 수 있다.

[0048] 플로우 가이드 메모리 모듈(120)은 공간 특징 엔코더 네트워크 모듈(100)로부터 출력되는 공간 특징맵 및 시간 특징 엔코더 네트워크 모듈(110)로부터 출력되는 시간 특징맵을 입력받아 신경망 연산을 통해 깊이 특징맵을 생성한다. 플로우 가이드 메모리 모듈(120)은 시간 특징맵 및 공간 특징맵을 순차적(t-1, t, t+1,쩜)으로 입력받아 깊이 특징맵을 생성한다.

[0049] 바람직하게는 시간 특징맵 및 공간 특징맵을 서로 결합(Concatenate)한 특징맵이 플로우 가이드 메모리 모듈

(120)로 입력된다. 특징맵간 결합을 위해 시간 특징맵의 차원과 깊이 특징맵의 차원은 동일한 것이 바람직하다.

[0050] 앞서 설명한 바와 같이, 연속하는 프레임들은 서로 독립적이지 않고 상관 관계가 있다. 본 발명은 이러한 상관 관계를 고려한 깊이 특징맵 생성을 위해 플로우 가이드 메모리 모듈(120)로 RNN(Recurrent Neural Network)를 사용한다. RNN 네트워크는 일반적인 CNN 네트워크와 비교하여 프레임간 상관 관계 또는 의존 관계를 보다 정확히 반영한 깊이 특징맵 생성이 가능하다.

[0051] RNN 네트워크는 다양한 종류의 네트워크를 포함한다. 기본적인 RNN 네트워크로 LSTM(Long Short-term Memory) 및 GRU(Gated Recurrent Unit)가 있다. 또한, 근래에는 콘볼루션을 LSTM 및 GRU에 각각 반영한 ConvLSTM 및 ConvGRU가 사용되기도 한다.

[0052] 본 발명의 바람직한 실시예에 따르면, RNN 네트워크 중 ConvGRU를 사용할 수 있다. RNN 네트워크 중 ConvGRU가 유리한 이유는 ConvGRU가 공간 해상도 손실이 크게 발생하지 않고 메모리 사용 관점에서 유리하기 때문이다. 물론, 다른 종류의 RNN 네트워크가 사용될 수도 있다는 점은 당업자에게 있어 자명할 것이다.

[0053] 플로우 가이드 메모리 모듈(120)의 상세한 동작 구조는 별도의 도면을 참조하여 추후 설명하기로 한다.

[0054] 플로우 가이드 메모리 모듈(120)로부터 출력되는 깊이 특징맵은 디코더 네트워크 모듈(130)로 입력된다. 디코더 네트워크 모듈(130)은 입력된 깊이 특징맵에 대한 신경망 연산을 통해 최종적인 깊이 맵을 생성한다. 디코더 네트워크 모듈(130)은 일례로 CNN을 이용하여 디코딩을 수행할 수 있을 것이나 이에 한정되는 것은 아니다. 디코더 네트워크 모듈(130)은 일반적인 디코더 네트워크와 같이 깊이 특징맵의 차원을 확장시키면서 깊이 맵을 생성할 수 있을 것이다.

[0055] 본 발명의 바람직한 실시예에 따르면, 엔코딩 과정에서의 특징 정보를 반영한 디코딩이 이루어질 수 있도록 공간 특징 엔코더 네트워크 모듈(100)에서 각 레이어별로 생성되는 특징맵들 및 시간 특징 엔코더 네트워크 모듈(110)에서 각 레이어별로 생성되는 특징맵들에 대한 스킵 커넥션(Skp Connection)이 이루어질 수도 있을 것이다. 스킵 커넥션은 디코딩 시 엔코딩 과정에서 생성된 특징맵들을 결합하여 다음 레이어의 디코딩에 사용하는 것으로서 스킵 커넥션은 다양한 뉴럴 네트워크의 엔코딩 및 디코딩에 사용되는 것이기에 이에 대한 상세한 설명은 생략하기로 한다.

[0056] 결국, 본 발명의 제1 실시예에 따른 깊이 추정 장치는 4개의 뉴럴 네트워크로 이루어져 있다고 할 수 있으며, 최종적인 깊이 맵은 디코더 네트워크 모듈(130)을 통해 출력되는 것이다.

[0057] 본 발명의 깊이 추정 장치를 구성하는 4개의 뉴럴 네트워크의 가중치 학습은 출력되는 깊이 맵에 대한 손실을 산출하고 이를 역전파하는 방식으로 이루어질 수 있을 것이다. 손실의 역전파는 역순으로 진행되어 디코더 네트워크 모듈(130) -> 플로우 가이드 메모리 모듈(120) -> 공간 특징 엔코더 네트워크 모듈(100)/시간 특징 엔코더 네트워크 모듈(110) 순서로 이루어지며, 손실을 최소화하기 위한 방향으로 가중치 업데이트가 진행된다.

[0058] 학습을 위한 손실 연산은 알려진 다양한 방법이 이용될 수 있으며, 본 발명에서 적용한 손실 연산에 대해서는 후에 설명하기로 한다.

[0059] 도 2는 본 발명의 제2 실시예에 따른 단안 비디오 영상의 깊이 추정을 위한 뉴럴 네트워크 구조를 나타낸 도면이다.

[0060] 본 발명의 제1 실시예에 따른 단안 비디오 영상의 깊이 추정 장치는 공간 특징 엔코더 네트워크 모듈(100), 시간 특징 엔코더 네트워크 모듈(110), 플로우 가이드 메모리 모듈(120), 디코더 네트워크 모듈(130) 및 광학 플로우 교정 네트워크 모듈(200)을 포함한다.

[0061] 본 발명의 제2 실시예에 따른 깊이 추정 장치는 제1 실시예와 비교할 때 광학 플로우 교정 네트워크 모듈(200)이 추가적으로 구비된다. 광학 플로우 교정 네트워크 모듈(200)을 제외한 다른 모듈들의 동작은 제1 실시예와 동일하다.

[0062] 광학 플로우 교정 네트워크 모듈(200)은 신경망 연산을 통해 교정된 시간 특징맵을 생성한다. 광학 플로우 영상만으로는 정확한 광학 플로우를 반영한 시간 특징맵을 생성하기 어렵다. 광학 플로우 교정 네트워크 모듈은 보자 정확한 시간 특징맵을 생성하기 위해 현재 프레임 영상(I^t), 이전 프레임 이미지(I^{t-1}) 및 광학 플로우 영상(O^t)을 입력받아 신경망 연산을 통해 교정된 시간 특징맵을 생성한다.

[0063] 교정된 시간 특징맵은 플로우 가이드 메모리 모듈(120)에서 이전 상태 특징맵의 워핑(Warping)에 이용된다. 이

전 상태 특징맵의 워핑에 대해서는 별도의 도면을 참조하여 상세히 설명한다.

[0064] 도 3은 본 발명의 일 실시예에 따른 광학 플로우 교정 네트워크 모듈의 동작 구조를 나타낸 도면이다.

[0065] 도 3을 참조하면, 광학 플로우 교정 네트워크 모듈(200)로는 현재 프레임 영상(I^t), 이전 프레임 영상(I^{t-1}) 및 광학 플로우 영상(O^t)이 서로 결합(Concatenate)된 형태로 입력된다.

[0066] 입력에 대한 콘볼루션 엔코딩을 통해 특징맵(Feature map)이 각 레이어별로 생성된다. 제1 레이어의 특징맵은 결합된 입력 영상과 동일한 사이즈를 가진다. 제1 레이어의 출력인 R^t_1은 제1 특징맵과 광학 플로우 영상(O^t)이 결합된 후 콘볼루션 엔코딩을 통해 획득된다.

[0067] 제2 레이어의 특징맵은 제1 레이어의 출력인 R^t_1와 제1 레이어 특징맵이 결합된 영상에 대한 콘볼루션 엔코딩을 통해 생성된다. 제2 레이어의 특징맵은 제1 레이어의 특징맵에 비해 1/2 다운샘플링된 사이즈를 가질 수 있다. 제2 레이어의 출력인 R^t_2는 제2 레이어 특징맵과 1/2 다운 샘플링된 광학 플로우 영상이 결합된 영상에 대한 콘볼루션 엔코딩을 통해 획득된다.

[0068] 제3 레이어의 특징맵은 제2 레이어의 출력인 R^t_2와 제2레이어 특징맵을 결합한 영상에 대한 콘볼루션 엔코딩을 통해 생성된다. 제3 레이어 특징맵은 제2 레이어 특징맵에 비해 1/2 다운 샘플링된 사이즈를 가질 수 있다.

[0069] 제3 레이어의 출력인 R^t_3는 제3 레이어 특징맵과 1/4 다운 샘플링된 광학 플로우 영상이 결합된 영상에 대한 콘볼루션 엔코딩을 통해 획득되며, R^t_3가 최종적으로 교정되는 시간 특징맵이고 이는 교정된 광학 플로우로 정의할 수도 있을 것이다.

[0070] 본 발명의 일 실시예에 따르면, 광학 플로우 교정 네트워크 모듈(200)에 대한 학습은 두 개의 손실을 이용하여 이루어질 수 있다. 제1 손실은 영상 일관성 손실(Photometric Consistency Loss)이고 제2 손실은 평활화 손실(Smoothness Loss)이다.

[0071] 영상 일관성 손실은 각 레이어에서 출력되는 광학 플로우(R^t_1, R^t_2, R^t_3)를 I^{t-1}에 적용한 영상 $\bar{I}_i^t = \mathcal{W}(I_i^{t-1}; R_i^t)$과 현재 프레임 영상($I^t$)과의 유사도로부터 연산되는 손실이다. 여기서 W()는 워핑 함수를 나타낸다.

[0072] 구체적으로, 영상 일관성 손실은 다음의 수학식 1과 같이 연산될 수 있다.

수학식 1

$$\mathcal{L}_i^{PH} = \frac{1}{N_i}\sum_p \left(\beta \frac{1 - \mathrm{SSIM}(I_i^t(p), \bar{I}_i^t(p))}{2} + (1-\beta)\|I_i^t(p) - \bar{I}_i^t(p)\|_1 \right)$$

[0073]

[0074] 위 수학식1에서, Ni는 모든 픽셀의 수이고, p는 픽셀을 나타내며, SSIM은 구조적 유사도를 연산하는 함수이고, 는 ß는 밸런스 상수이고 0 내지 1의 값 중 하나로 선택된다.

[0075] 제2 손실인 평활화(Smoothness) 손실은 광학 플로우 영상의 평활화 정도를 연산하는 것으로서, 영상 분야에서 학습을 위해 일반적으로 사용하는 손실이다. 일례로, 평활화 손실은 다음의 수학식 2와 같이 연산될 수 있다.

수학식 2

$$\mathcal{L}_i^{OS} = \frac{1}{N_i} \sum_p \|\nabla^2 R_i^t(p)\|_1 \; e^{-\gamma \|\nabla^2 I_i^t(p)\|_1}$$

[0076]

[0077] 위 수학식에서, τ는 임의로 설정되는 상수이다.

[0078] 물론, 광학 플로우 교정 네트워크 모듈(200)의 학습은 위에서 설명한 광학 일관성 손실 및 평활화 손실 이외에도 다양한 방식으로 학습될 수도 있을 것이며, 이러한 학습 방식의 변경이 본 발명의 사상에 영향을 미치지 않는다는 점을 당업자라면 이해할 수 있을 것이다.

[0079] 결국, 제1 실시예와 제2 실시예의 차이는 플로우 가이드 메모리 모듈(120)에서의 이전 상태 특징맵 워핑 시 시간 특징 엔코더 네트워크 모듈(110)에서 출력되는 시간 특징맵을 이용할지 아니면 광학 플로우 교정 네트워크 모듈(200)에서 출력되는 교정된 시간 특징맵을 이용할지 여부이다.

[0080] 제1 실시예는 별도의 교정된 시간 특징맵을 획득하지 않고 시간 특징 엔코더 네트워크 모듈(110)에서 획득된 시간 특징맵을 이용하여 플로우 가이드 메모리 모듈(120)에서 이전 상태 특징맵을 워핑한다. 그러나, 제2 실시예는 플로우 가이드 메모리 모듈(120)에서의 이전 상태 특징맵의 워핑에 획득된 시간 특징맵을 이용하지 않고 교정된 광학 플로우를 이용하는 것이다.

[0082] 다만, 플로우 가이드 메모리 모듈의 입력으로는 제2 실시예에서도 교정된 시간 특징맵이 입력되지 아니하고 시간 특징 엔코더 네트워크 모듈(110)로부터 출력되는 시간 특징맵이 입력된다.

[0083] 도 4는 본 발명의 일 실시예에 따른 플로우 가이드 메모리 모듈의 동작 구조를 나타낸 도면이다.

[0084] 도 4는 본 발명의 일 실시예에 따른 플로우 가이드 메모리 모듈(120)이 ConvGRU를 사용하는 경우를 예로 한 동작 구조가 도시되어 있다. 그러나, 앞서 설명한 바와 같이 ConvGRU 이외에도 다양한 RNN 네트워크가 사용될 수도 있을 것이다.

[0085] ConvGRU에는 5개의 값이 사용된다. h^t는 현재 상태 특징맵이고, h^{t-1}은 이전 상태 특징맵이며, r^t는 리셋 게이트, z^t는 업데이트 게이트이며, $\tilde{h}^t$는 후보 상태 특징맵이다.

[0086] 종래의 ConvGRU에서 현재 상태 특징맵, 이전 상태 특징맵, 리셋 게이트 , 업데이트 게이트 및 후보 상태 특징맵은 다음의 수학식 3과 같이 연산된다.

수학식 3

$$z^t = \sigma(W_{xz} * x^t + W_{hz} * h^{t-1} + b_z)$$
$$r^t = \sigma(W_{xr} * x^t + W_{hr} * h^{t-1} + b_r)$$
$$\tilde{h^t} = \tanh(W_{x\tilde{h}} * x^t + r^t \odot (W_{h\tilde{h}} * h^{t-1}) + b_{\tilde{h}})$$

$$h^t = (1 - z^t) \odot h^{t-1} + z^t \odot \tilde{h^t},$$

[0087]

[0088] 위 수학식 3에서, σ는 시그모이드 함수를 의미하고, $\odot$는 엘리먼트-와이즈(element-wise) 곱셉을 의미하며, *는 콘볼루션을 의미하며, x^t는 입력되는 특징맵을 의미한다. 본 발명에서는 공간 특징맵과 시간 특징맵을 결합한 특징맵이 x^t가 된다.

[0089] 앞서 설명한 바와 같이, 종래의 ConvGRU를 이용하게 될 경우 프레임 시간 간격 사이의 움직임을 반영하기 어렵게 된다. 이전 프레임과 현재 프레임이 서로 상관 관계에 있고, 이전 프레임과 현재 프레임의 깊이가 서로 상관

관계에 있기는 하나, 깊이의 상관 관계는 현재 프레임과 이전 프레임이 동일한 상태에 있을 때 보다 정확한 상관 관계가 획득될 수 있다. 이에, 현재 프레임과 이전 프레임 사이에 객체의 이동 또는 카메라의 이동이 있을 경우 이를 보상한 상태에서 깊이의 상관 관계를 고려하는 것이 바람직하다.

[0090] 본 발명에서는 획득되는 시간 특징맵 또는 교정된 시간 특징맵을 이용하여 워핑을 통해 이전 상태 특징맵을 보정하고, 이전 상태 특징맵 대신 보정된 보정 상태 특징맵을 이용한다. 보정 상태 특징맵은 도 4에서 $\bar{h}^t$로 정의한다.

[0091] 결국, 본 발명의 플로우 가이드 메모리 모듈(120)은 이전 상태 특징맵을 시간 특징맵 또는 교정된 시간 특징맵에 기초하여 보정한 후 보정 상태 특징맵을 적용하여 현재 상태 특징맵을 생성하는 것이다.

[0092] 본 발명에 따른 현재 상태 특징맵, 보정 상태 특징맵, 리셋 게이트, 업데이트 게이트 및 후보 상태 특징맵은 다음의 수학식 4와 같이 연산된다.

수학식 4

$$\bar{h}^t = M^t \odot \mathcal{W}(h^{t-1}; R_3^t)$$
$$z^t = \sigma(W_{xz} * x^t + W_{\bar{h}z} * \bar{h}^t + b_z)$$
$$r^t = \sigma(W_{xr} * x^t + W_{\bar{h}r} * \bar{h}^t + b_r)$$
$$\tilde{h}^t = \tanh(W_{x\tilde{h}} * x^t + r^t \odot (W_{\bar{h}\tilde{h}} * \bar{h}^t) + b_{\tilde{h}})$$

$$h^t = (1 - z^t) \odot \bar{h}^t + z^t \odot \tilde{h}^t,$$

[0093]

[0094] 수학식 4에서, $\mathcal{W}(h^{t-1}; R_3^t)$는 시간 특징맵 또는 교정된 시간 특징맵을 이용하여 이전 상태 특징맵을 워핑한 특징맵이고 $\mathcal{W}(h^{t-1}; R_3^t)(p) = h^{t-1}(p + R_3^t(p))$로 정의될 수 있다. 또한, W는 미리 설정되는 가중치이고, b는 미리 설정되는 바이어스 값이다.

[0095] 한편, 위 수학식 4에서 워핑된 특징맵에는 마스크 특징맵 M^t가 적용되어 있다. 마스크 특징맵은 각 픽셀(P)별로 워핑의 신뢰도를 나타내는 특징맵으로 정의할 수 있다. 만일, 현재 프레임 영상($I^t_3(p)$)과 시간 특징맵(광학 플로우)에 의해 워핑된 이전 프레임 영상($\bar{I}^t_3(p)$)과의 차이가 크다면 워핑의 신뢰도가 높지 않을 것이고 차이가 크지 않으면 워핑의 신뢰도가 높을 것이다. 이러한 사실에 기초하여, 마스크 특징맵 M^t는 다음의 수학식 5와 같이 정의될 수 있을 것이다.

수학식 5

$$M^t(p) = \exp(-\epsilon \| I_3^t(p) - \bar{I}_3^t(p) \|_1)$$

[0096]

[0097] 위 수학식 5에서, ε은 임의로 설정되는 상수이며, ε에 의해 지수 함수의 폭이 결정된다.

[0098] 도 5는 본 발명의 일 실시예에 따른 플로우 가이드 메모리 모듈에서의 워핑을 개념적으로 나타낸 도면이다.

[0099] 도 5를 참조하면, (t-1) 프레임에서 t 프레임 사이에 자동차가 이동하는 경우가 도시되어 있다. 기존 ConvGRU 방식에 의할 경우 이전 상태 특징맵인 h^{t-1}이 이용되기에 자동차의 움직임이 반영된 깊이의 상관 관계가 깊이 추정에 이용되기 어렵다.

[0100] 본 발명은 이러한 문제를 해결하기 위해 광학 플로우(시간 특징맵)를 이용하여 이전 상태 특징맵 ht-1을 $\bar{h}^t$로 보정하고, 보정 상태 특징맵을 이용하여 깊이 특징맵을 연산하도록 하는 것이다.

[0101] 한편, 도 1에 도시된 본 발명의 일 실시예에 따른 깊이 추정 장치를 구성하는 뉴럴 네트워크들의 학습을 위해 깊이맵 참값($G^t(p)$)과 디코더 네트워크 모듈을 통해 출력된 깊이맵과의 차에 상응하는 손실이 이용될 수 있다.

[0102] 보다 정확한 학습을 위해 참값과 차이에 대한 손실(L^D)과 평활화 손실(L^{DS})이 함께 이용될 수 있다.

[0103] 본 발명의 일 실시예에 따르면, 깊이맵 참값과 차이에 대한 손실은 다음의 수학식 6과 같이 연산될 수 있다.

수학식 6

[0104]
$$\mathcal{L}^{SI} = \frac{1}{N}\sum_{p} s^2(p) - \frac{\alpha}{N^2}\sum_{p,q} s(p)s(q)$$

[0105] 위 수학식 6에서 $s(p) = \log D^t(p) - \log G^t(p)$ 로 정의되고, $D^t(p)$는 출력된 깊이맵이고, $G^t(p)$는 참값(Ground Truth) 깊이맵이다.

[0106] 위 수학식 7에서, 첫번째 텀은 출력된 깊이맵과 참값 깊이맵 사이의 차를 의미한다. 그런데, 단안 비디오 영상 시퀀스에서 각 픽셀의 참값 깊이를 획득하는 것은 매우 어렵다. 수학식 7의 두번째 텀은 이러한 문제를 완화시키기 위한 텀이다. 두 개의 픽셀 페어인 p와 q에 대해 s(p)와 s(q)의 곱이 합해지며, α는 0 내지 1의 값을 가지는 밸런스 상수이고 N은 모든 픽셀의 수이다.

[0107] 또한, 평활화 손실은 깊이의 불연속을 방지하기 위해 연산되며, 다음의 수학식 7과 같이 연산될 수 있을 것이다.

수학식 7

[0108]
$$\mathcal{L}^{DS} = \frac{1}{N}\sum \|\nabla^2 D^t(p)\|_1 \; e^{-\gamma\|\nabla^2 I^t(p)\|_1}$$

[0109] 위의 설명된 예에서, 도 1에 도시된 깊이 추정 장치의 학습을 위해 역전파되는 손실은 참값과 차이에 대한 손실(L^D)과 평활화 손실(L^{DS})의 합으로 연산된다.

[0110] 도 6은 본 발명의 제2 실시예에 따른 단안 비디오 영상의 깊이 추정 방법의 전체적인 흐름을 도시한 순서도이다.

[0111] 도 6을 참조하면, 우선 현재 프레임 영상을 공간 특징 엔코더 네트워크 모듈(100)에 입력하여 공간 특징맵을 생성한다(단계 600).

[0112] 또한, 현재 프레임 영상과 이전 프레임 영상을 이용하여 획득되는 광학 플로우 영상을 시간 특징 엔코더 네트워크 모듈(110)에 입력하여 시간 특징맵을 생성한다(단계 602).

[0113] 한편, 현재 프레임 영상, 이전 프레임 영상 및 광학 플로우 영상을 광학 플로우 교정 네트워크 모듈(200)에 입력하여 교정된 시간 특징맵을 생성한다(단계 604).

[0114] 단계 600에서 생성되는 공간 특징맵 및 단계 602에서 시간 특징맵은 서로 결합되어 광학 플로우 메모리 모듈(120)로 입력되며, 플로우 가이드 메모리 모듈(120)은 신경망 연산을 통해 깊이 특징맵을 생성한다(단계 606). 광학 플로우 메모리 모듈은 RNN을 이용한다. 플로우 가이드 메모리 모듈(120)은 RNN에서 현재 상태 특징맵의 갱신에 이용되는 이전 상태 특징맵을 단계 604에서 생성된 교정된 시간 특징맵의 광학 플로우를 이용하여 워핑함으로써 보정 상태 특징맵을 생성하고, 보정 상태 특징맵을 현재 상태 특징맵의 갱신에 이용한다.

[0115] 단계 606에서 생성되는 깊이 특징맵은 디코더 네트워크 모듈(130)로 입력되며, 디코더 네트워크 모듈(130)은 신경망 연산을 통해 깊이 맵을 생성한다(단계 608).

[0116] 한편, 도 6에서는 제2 실시예의 경우를 예로 하여 전체적인 흐름을 설명하였으나, 제1 실시예는 시간 특징 엔코더 네트워크 모듈로부터 출력되는 시간 특징맵을 RNN의 이전 상태 특징맵의 워핑에 사용한다는 점에서만 차이가 있다는 점을 위에서 설명하였다.

[0117] 전술한 본 발명의 설명은 예시를 위한 것이며, 본 발명이 속하는 기술분야의 통상의 지식을 가진 자는 본 발명의 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해할 수 있을 것이다.

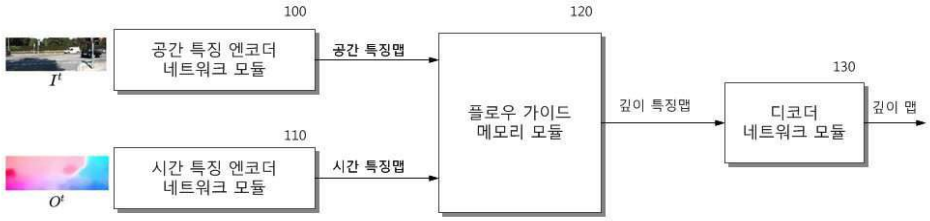
[0118] 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다.

[0119] 예를 들어, 단일형으로 설명되어 있는 각 구성 요소는 분산되어 실시될 수도 있으며, 마찬가지로 분산된 것으로 설명되어 있는 구성 요소들도 결합된 형태로 실시될 수 있다.

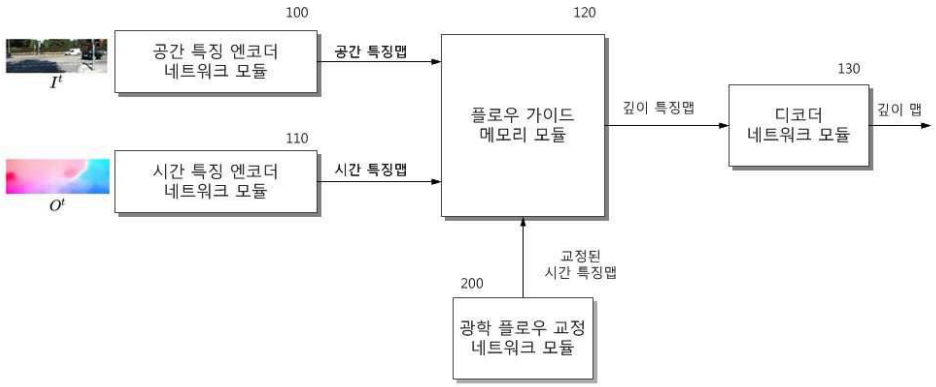
[0120] 본 발명의 범위는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

도면

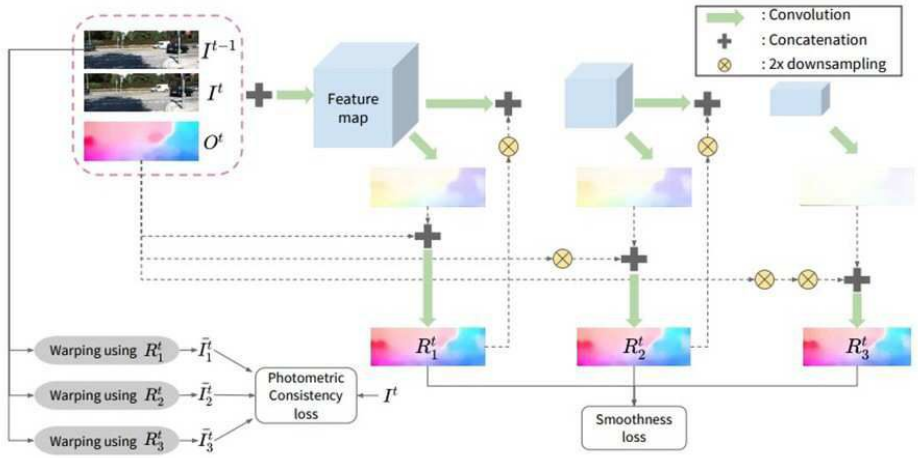
도면1



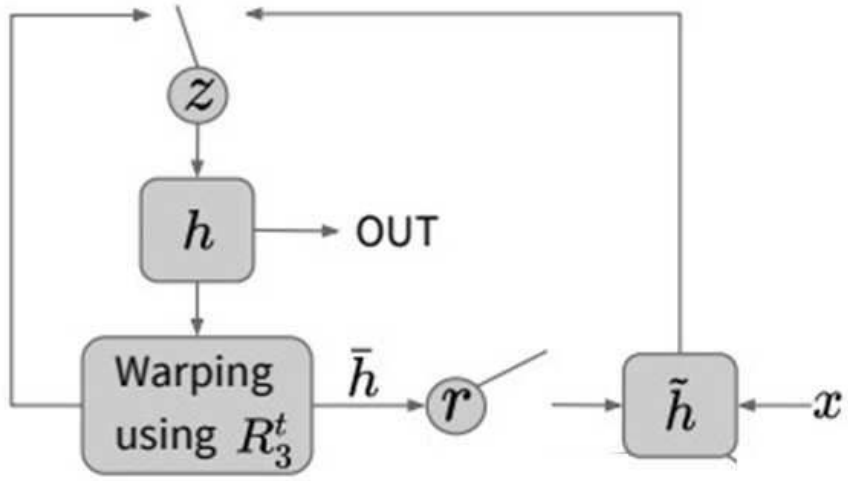
도면2



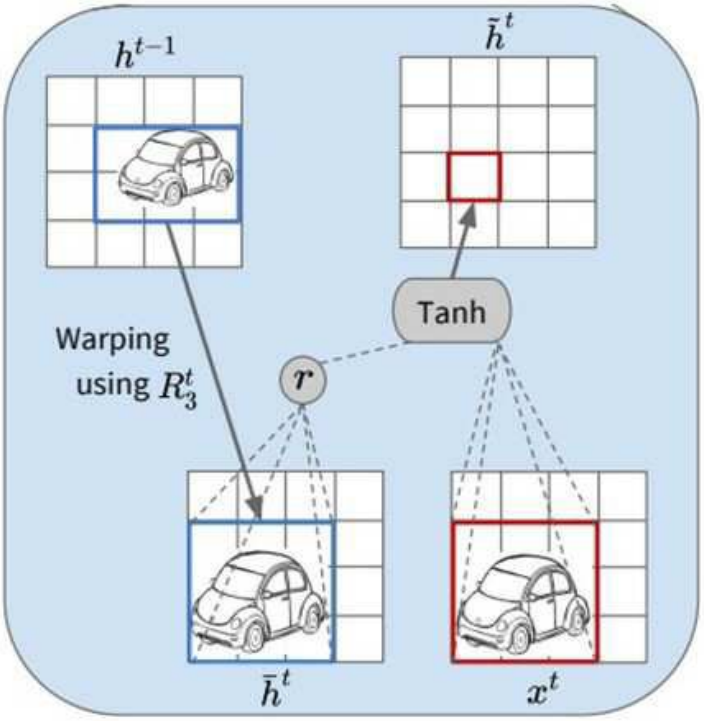
도면3


: Convolution
: Concatenation
: 2x downsampling
Feature map
Warping using
Photometric Consistency loss
Smoothness loss

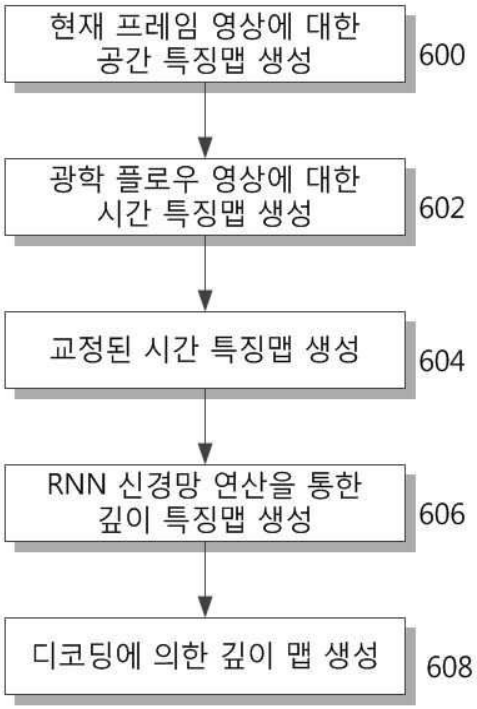
도면4


OUT
Warping using

도면5



도면6



【심사관 직권보정사항】

【직권보정 1】

【보정항목】 청구범위

【보정세부항목】 청구항 5

【변경전】

제1항에 있어서,

상기 신뢰도는 상기 시간 특징맵 또는 상기 교정된 시간 특징맵에 기초하여 이전 프레임 영상을 워핑한 영상과 현재 프레임 영상간의 차에 기초하여 연산되는 것을 특징으로 하는 단안 비디오 영상의 깊이 추정 장치.

【변경후】

제4항에 있어서,

상기 신뢰도는 상기 시간 특징맵 또는 상기 교정된 시간 특징맵에 기초하여 이전 프레임 영상을 워핑한 영상과 현재 프레임 영상간의 차에 기초하여 연산되는 것을 특징으로 하는 단안 비디오 영상의 깊이 추정 장치.

【직권보정 2】

【보정항목】청구범위

【보정세부항목】청구항 13

【변경전】

제9항에 있어서,

상기 신뢰도는 상기 시간 특징맵 또는 상기 교정된 시간 특징맵에 기초하여 이전 프레임 영상을 워핑한 영상과 현재 프레임 영상간의 차에 기초하여 연산되는 것을 특징으로 하는 단안 비디오 영상의 깊이 추정 방법.

【변경후】

제12항에 있어서,

상기 신뢰도는 상기 시간 특징맵 또는 상기 교정된 시간 특징맵에 기초하여 이전 프레임 영상을 워핑한 영상과 현재 프레임 영상간의 차에 기초하여 연산되는 것을 특징으로 하는 단안 비디오 영상의 깊이 추정 방법.