(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0082901
(43) 공개일자 2021년07월06일

(51) 국제특허분류(Int. Cl.)
G06N 3/08 (2006.01) *G06N 3/04* (2006.01)
H04N 9/67 (2006.01)
(52) CPC특허분류
G06N 3/08 (2013.01)
G06N 3/0454 (2013.01)
(21) 출원번호 10-2019-0175328
(22) 출원일자 2019년12월26일
심사청구일자 2019년12월26일

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
고성제
서울특별시 서초구 방배로 270, 나-304호 (방배동, 삼호아파트)
엄광현
서울특별시 종로구 홍지문길 69, 105동 204호 (구기동, 청구빌라)
조성진
경기도 성남시 수정구 위례중앙로 216, 5313동 1102호 (창곡동, 위례 호반베르디움)
(74) 대리인
윤귀상

전체 청구항 수 : 총 13 항

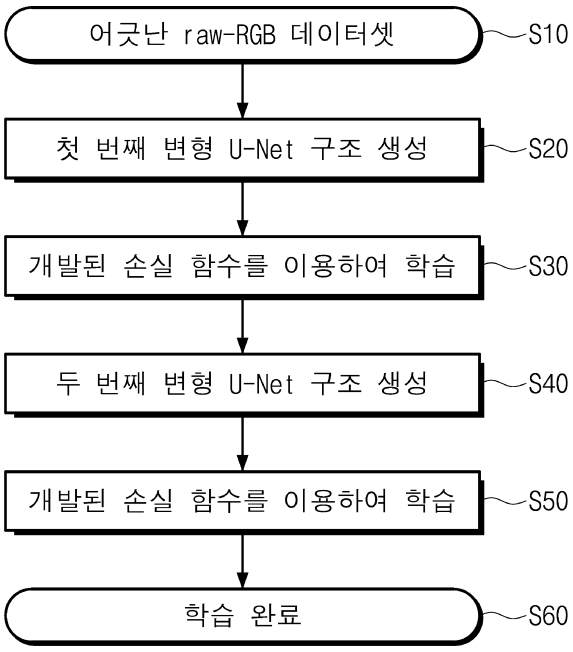
(54) 발명의 명칭 **두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 방법, 이를 수행하기 위한 기록 매체 및 장치**

(57) 요 약

두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 방법은, 서로 다른 카메라로부터 입력 raw 영상과 출력 RGB 영상의 어긋난 데이터 세트를 획득하는 단계; 다수의 풀링(Pooling) 계층과 다수의 업샘플링(Upsampling) 계층으로 구성된 제1 U-Net 신경망 모델을 생성하는 단계; 출력 영상과 타겟 영상 사이의 픽셀

(뒷면에 계속)

대 표 도 - 도6



손실, 특징 손실 및 색상 손실의 합으로 정의되는 손실 함수를 이용하여, 상기 데이터 세트를 상기 제1 U-Net 신경망 모델을 통해 학습하는 단계; 다수의 폴링 계층과 다수의 업샘플링 계층으로 구성된 제2 U-Net 신경망 모델을 생성하는 단계; 및 상기 손실 함수를 이용하여, 상기 제1 U-Net 신경망 모델을 통해 학습된 결과를 상기 제2 U-Net 신경망 모델 학습을 통해 학습하여 raw 영상과 RGB 영상의 데이터 쌍(pair)으로 정렬하는 단계;를 포함한다. 이에 따라, raw 센서로부터 획득한 영상을 입력으로 사용하여 DSLR 수준의 화질의 RGB 영상을 생성할 수 있도록 합성곱 신경망 모델을 학습할 수 있다.

(52) CPC특허분류
H04N 9/67 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711081163
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기술진흥센터
연구사업명	인공지능
연구과제명	(딥뷰-2세부) 대규모 실시간 비디오 분석에 의한 전역적 다중 관심객체 추적 및 상황 예측 기술 개발
기 여 율	1/1
과제수행기관명	광주과학기술원
연구기간	2019.01.01 ~ 2019.12.31

명 세 서

청구범위

청구항 1

서로 다른 카메라로부터 입력 raw 영상과 출력 RGB 영상의 어긋난 데이터 세트를 획득하는 단계;

다수의 풀링(Pooling) 계층과 다수의 업샘플링(Upsampling) 계층으로 구성된 제1 U-Net 신경망 모델을 생성하는 단계;

출력 영상과 타겟 영상 사이의 픽셀 손실, 특징 손실 및 색상 손실의 합으로 정의되는 손실 함수를 이용하여, 상기 데이터 세트를 상기 제1 U-Net 신경망 모델을 통해 학습하는 단계;

다수의 풀링 계층과 다수의 업샘플링 계층으로 구성된 제2 U-Net 신경망 모델을 생성하는 단계; 및

상기 손실 함수를 이용하여, 상기 제1 U-Net 신경망 모델을 통해 학습된 결과를 상기 제2 U-Net 신경망 모델 학습을 통해 학습하여 raw 영상과 RGB 영상의 데이터 쌍(pair)으로 정렬하는 단계;를 포함하는, 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 방법.

청구항 2

제1항에 있어서, 상기 제1 U-Net 신경망 모델을 통해 학습하는 단계는,

상기 데이터 세트를 출력 영상과 타겟 영상의 차이를 이용한 상기 손실 함수를 정의하는 단계를 포함하는, 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 방법.

청구항 3

제2항에 있어서, 상기 손실 함수를 정의하는 단계는,

상기 데이터 세트를 출력 영상과 타겟 영상의 픽셀 차이인 픽셀 손실을 계산하는 단계;

미리 설정된 레벨 이상의 특징들에 기반한 퍼셉츄얼(perceptual) 손실로부터 특징을 추출하는 특징 손실을 계산하는 단계; 및

RGB 색상 벡터 사이의 코사인 거리로부터 색상 손실을 계산하는 단계;를 포함하는, 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 방법.

청구항 4

제3항에 있어서, 상기 미리 설정된 레벨 이상의 특징들에 기반한 퍼셉츄얼 손실로부터 특징을 추출하는 특징 손실을 계산하는 단계는,

ReLU(Rectified Linear Unit) 4 이상의 계층에서 특징을 추출하는, 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 방법.

청구항 5

제3항에 있어서, 상기 RGB 색상 벡터 사이의 코사인 거리로부터 색상 손실을 계산하는 단계는,

영상을 2배 축소하여 사용하는, 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 방법.

청구항 6

제1항에 있어서,

학습된 raw 영상과 RGB 영상의 데이터 쌍(pair)을 기초로, 실시간으로 raw 영상으로부터 RGB 영상을 출력하는 단계를 더 포함하는, 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 방법.

청구항 7

제1항에 있어서,

상기 제1 U-Net 신경망 모델 및 상기 제2 U-Net 신경망 모델은, 채널 어텐션 메커니즘(channel attention mechanism)과 롱 스킵 커넥션(long skip connection)을 포함하는, 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 방법.

청구항 8

제1항 내지 제7항의 어느 하나의 항에 있어서,

상기 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 방법을 수행하기 위한 컴퓨터 프로그램이 기록된 컴퓨터로 판독 가능한 저장 매체.

청구항 9

다수의 풀링(Pooling) 계층과 다수의 업샘플링(Upsampling) 계층으로 구성된 제1 U-Net 신경망 모델;

다수의 풀링(Pooling) 계층과 다수의 업샘플링(Upsampling) 계층으로 구성되고, 상기 제1 U-Net 신경망 모델과 연속되어 W-Net 구조를 형성하는 제2 U-Net 신경망 모델;

출력 영상과 타겟 영상 사이의 픽셀 손실, 특징 손실 및 색상 손실의 합으로 정의되는 손실 함수를 정의하는 손실 함수부; 및

상기 손실 함수를 이용하여 상기 제1 U-Net 신경망 모델 및 상기 제2 U-Net 신경망 모델을 통해 서로 다른 카메라로부터 입력 raw 영상과 출력 RGB 영상의 어긋난 데이터 세트를 학습하여, raw 영상과 RGB 영상의 데이터 쌍(pair)으로 정렬하는 영상 학습부;를 포함하는, 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 장치.

청구항 10

제9항에 있어서, 상기 손실 함수부는,

상기 데이터 세트를 출력 영상과 타겟 영상의 픽셀 차이인 픽셀 손실을 계산하는 픽셀 손실부;

미리 설정된 레벨 이상의 특징들에 기반한 퍼셉츄얼(perceptual) 손실로부터 특징을 추출하는 특징 손실을 계산하는 특징 손실부; 및

RGB 색상 벡터 사이의 코사인 거리로부터 색상 손실을 계산하는 색상 손실부;를 포함하는, 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 장치.

청구항 11

제10항에 있어서,

상기 특징 손실부는 ReLU(Rectified Linear Unit) 4 이상의 계층에서 특징을 추출하고, 상기 색상 손실부는 영상을 2배 축소하여 사용하는, 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 장치.

청구항 12

제9항에 있어서,

학습된 raw 영상과 RGB 영상의 데이터 쌍(pair)을 기초로, 실시간으로 raw 영상으로부터 RGB 영상을 출력하는 영상 변환부를 더 포함하는, 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 장치.

청구항 13

제9항에 있어서,

상기 제1 U-Net 신경망 모델 및 상기 제2 U-Net 신경망 모델은, 채널 어텐션 메커니즘(channel attention mechanism)과 롱 스킵 커넥션(long skip connection)을 포함하는, 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 장치.

발명의 설명

기 술 분 야

[0001] 본 발명은 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 방법, 이를 수행하기 위한 기록 매체 및 장치에 관한 것으로서, 더욱 상세하게는 딥 러닝을 기반으로 스마트폰 카메라로부터 얻어진 raw Bayer 영상을 DSLR 카메라로 획득한 RGB 영상으로 매핑하는 기법에 관한 것이다.

[0003] [국가지원 연구개발에 대한 설명]

[0004] 본 연구는 2019년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구이다 (No. 2014-3-00077, 대규모 실시간 비디오 분석에 의한 전역적 다중 관심객체 추적 및 상황 예측 기술 개발).

배 경 기 술

[0006] 스마트폰 카메라가 급격히 발전하고 있지만, 물리적인 한계(작은 센서 크기, 작은 렌즈 크기, 하드웨어 비용 등)로 인하여 DSLR 카메라로 촬영한 영상보다 낮은 화질의 영상을 얻게 된다.

[0007] 따라서, 스마트폰 카메라로 촬영한 영상의 화질을 개선하는 연구가 최근 진행되고 있다. 일부 스마트폰은 카메라 센서에서 획득한 raw 영상 데이터를 저장하는 것을 지원한다. 이러한 raw 영상은 카메라에 탑재된 영상 신호 처리 프로세서(Image Signal Processor)를 통해 RGB 영상으로 변환된다.

[0008] 그러나, 종래 기술들은 모바일 카메라 영상을 DSLR 카메라 수준으로 화질 개선을 하기 위해 RGB 영상만을 이용하였다. 이에 따라, RGB 영상은 이미 카메라에 탑재된 영상 신호 처리기(Image signal processor, ISP)를 통하여 raw 센서 데이터를 가공한 것으로, 데이터의 손실이 발생하는 문제점이 있다.

선행기술문헌

특허문헌

[0010] (특허문헌 0001) US 8,508,612 B2

(특허문헌 0002) KR 10-1996730 B1

(특허문헌 0003) US 2017/0270122 A1

비특허문헌

[0011] (비특허문헌 0001) D. G. Lowe, "Distinctive image features from scale-invariant keypoints," in International Journal of Computer Vision (IJCV), 60(2):91-110, 2004.

(비특허문헌 0002) E. Schwartz, R. Giryes and A. M. Bronstein, "DeepISP: Toward Learning an End-to-End Image Processing Pipeline," in IEEE Transactions on Image Processing, vol. 28, no. 2, pp. 912-923, Feb. 2019.

발명의 내용

해결하려는 과제

[0012] 이에, 본 발명의 기술적 과제는 이러한 점에서 착안된 것으로 본 발명의 목적은 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 방법을 제공하는 것이다.

[0013] 본 발명의 다른 목적은 상기 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 방법을 수행하기 위한 컴퓨터 프로그램이 기록된 기록 매체를 제공하는 것이다.

[0014] 본 발명의 또 다른 목적은 상기 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 방법을 수행하기 위한 장치를 제공하는 것이다.

과제의 해결 수단

[0016] 상기한 본 발명의 목적을 실현하기 위한 일 실시예에 따른 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 방법은, 서로 다른 카메라로부터 입력 raw 영상과 출력 RGB 영상의 어긋난 데이터 세트를 획득하는 단계; 다수의 풀링(Pooling) 계층과 다수의 업샘플링(Upsampling) 계층으로 구성된 제1 U-Net 신경망 모델을 생성하는 단계; 출력 영상과 타겟 영상 사이의 픽셀 손실, 특징 손실 및 색상 손실의 합으로 정의되는 손실 함수를 이용하여, 상기 데이터 세트를 상기 제1 U-Net 신경망 모델을 통해 학습하는 단계; 다수의 풀링 계층과 다수의 업샘플링 계층으로 구성된 제2 U-Net 신경망 모델을 생성하는 단계; 및 상기 손실 함수를 이용하여, 상기 제1 U-Net 신경망 모델을 통해 학습된 결과를 상기 제2 U-Net 신경망 모델 학습을 통해 학습하여 raw 영상과 RGB 영상의 데이터 쌍(pair)으로 정렬하는 단계;를 포함한다.

[0017] 본 발명의 실시예에서, 상기 제1 U-Net 신경망 모델을 통해 학습하는 단계는, 상기 데이터 세트를 출력 영상과 타겟 영상의 차이를 이용한 상기 손실 함수를 정의하는 단계를 포함할 수 있다.

[0018] 본 발명의 실시예에서, 상기 손실 함수를 정의하는 단계는, 상기 데이터 세트를 출력 영상과 타겟 영상의 픽셀 차이인 픽셀 손실을 계산하는 단계; 미리 설정된 레벨 이상의 특징들에 기반한 퍼셉츄얼(perceptual) 손실로부터 특징을 추출하는 특징 손실을 계산하는 단계; 및 RGB 색상 벡터 사이의 코사인 거리로부터 색상 손실을 계산하는 단계;를 포함할 수 있다.

[0019] 본 발명의 실시예에서, 상기 미리 설정된 레벨 이상의 특징들에 기반한 퍼셉츄얼 손실로부터 특징을 추출하는 특징 손실을 계산하는 단계는, ReLU(Rectified Linear Unit) 4 이상의 계층에서 특징을 추출할 수 있다.

[0020] 본 발명의 실시예에서, 상기 RGB 색상 벡터 사이의 코사인 거리로부터 색상 손실을 계산하는 단계는, 영상을 2배 축소하여 사용할 수 있다.

[0021] 본 발명의 실시예에서, 상기 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 방법은, 학습된 raw 영상과 RGB 영상의 데이터 쌍(pair)을 기초로, 실시간으로 raw 영상으로부터 RGB 영상을 출력하는 단계를 더 포함할 수 있다.

[0022] 본 발명의 실시예에서, 상기 제1 U-Net 신경망 모델 및 상기 제2 U-Net 신경망 모델은 채널 어텐션 메커니즘(channel attention mechanism)과 롱 스킵 커넥션(long skip connection)을 포함할 수 있다.

[0023] 상기한 본 발명의 다른 목적을 실현하기 위한 일 실시예에 따른 컴퓨터로 판독 가능한 저장 매체에는, 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 방법을 수행하기 위한 컴퓨터 프로그램이 기록되어 있다.

[0024] 상기한 본 발명의 또 다른 목적을 실현하기 위한 일 실시예에 따른 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 장치는, 다수의 풀링(Pooling) 계층과 다수의 업샘플링(Upsampling) 계층으로 구성된 제1 U-Net 신경망 모델; 다수의 풀링(Pooling) 계층과 다수의 업샘플링(Upsampling) 계층으로 구성되고, 상기 제1 U-Net 신경망 모델과 연속되어 W-Net 구조를 형성하는 제2 U-Net 신경망 모델; 출력 영상과 타겟 영상 사이의 픽셀 손실, 특징 손실 및 색상 손실의 합으로 정의되는 손실 함수를 정의하는 손실 함수부; 및 상기 손실 함수를 이용하여 상기 제1 U-Net 신경망 모델 및 상기 제2 U-Net 신경망 모델을 통해 서로 다른 카메라로부터 입력 raw 영상과 출력 RGB 영상의 어긋난 데이터 세트를 학습하여, raw 영상과 RGB 영상의 데이터 쌍(pair)으로 정렬하는 영상 학습부;를 포함한다.

[0025] 본 발명의 실시예에서, 상기 손실 함수부는, 상기 데이터 세트를 출력 영상과 타겟 영상의 픽셀 차이인 픽셀 손실을 계산하는 픽셀 손실부; 미리 설정된 레벨 이상의 특징들에 기반한 퍼셉츄얼(perceptual) 손실로부터 특징을 추출하는 특징 손실을 계산하는 특징 손실부; 및 RGB 색상 벡터 사이의 코사인 거리로부터 색상 손실을 계산하는 색상 손실부;를 포함할 수 있다.

[0026] 본 발명의 실시예에서, 상기 특징 손실부는 ReLU(Rectified Linear Unit) 4 이상의 계층에서 특징을 추출하고, 상기 색상 손실부는 영상을 2배 축소하여 사용할 수 있다.

[0027] 본 발명의 실시예에서, 상기 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 장치는, 학습된 raw 영상과 RGB 영상의 데이터 쌍(pair)을 기초로, 실시간으로 raw 영상으로부터 RGB 영상을 출력하는 영상 변환부를 더 포함할 수 있다.

[0028] 본 발명의 실시예에서, 상기 제1 U-Net 신경망 모델 및 상기 제2 U-Net 신경망 모델은 채널 어텐션 메커니즘(channel attention mechanism)과 롱 스킵 커넥션(long skip connection)을 포함할 수 있다.

발명의 효과

[0030] 이와 같은 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 방법에 따르면, 동일한 씬을 서로 다른 카메라로 촬영하여 U-Net 신경망 모델을 학습시킨 후 출력에 추가적으로 U-Net 신경망 모델을 적용하여 학습 시킨다. 이에 따라, 스마트폰으로 촬영된 영상을 DSLR 촬영 영상 정도의 퀄리티로 변환시킬 수 있다.

도면의 간단한 설명

[0032] 도 1은 본 발명의 일 실시예에 따른 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 장치의 블록도이다.

도 2는 어긋난 raw 영상과 RGB 영상의 예시를 보여주는 도면이다.

도 3은 본 발명의 사용하는 하나의 신경망 구조를 보여주는 개념도이다.

도 4는 본 발명의 제1 U-Net 신경망 모델 및 제2 U-Net 신경망 모델이 형성하는 두 단계 W-Net 구조를 보여주는 개념도이다.

도 5는 도 1의 손실 함수부의 상세한 블록도이다.

도 6은 본 발명의 일 실시예에 따른 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 방법의 흐름도이다.

도 7은 본 발명의 네트워크 구조 향상에 따른 성능 비교를 보여주는 표이다.

도 8은 본 발명의 손실 함수 조합에 따른 성능 비교를 보여주는 표이다.

도 9는 본 발명의 데이터 셋에서의 성능 비교를 보여주는 표이다.

도 10은 본 발명에 따른 결과의 예시를 보여주는 도면이다.

발명을 실시하기 위한 구체적인 내용

[0033] 후술하는 본 발명에 대한 상세한 설명은, 본 발명이 실시될 수 있는 특정 실시예를 예시로서 도시하는 첨부 도면을 참조한다. 이들 실시예는 당업자가 본 발명을 실시할 수 있기에 충분하도록 상세히 설명된다. 본 발명의 다양한 실시예는 서로 다르지만 상호 배타적일 필요는 없음이 이해되어야 한다. 예를 들어, 여기에 기재되어 있는 특정 형상, 구조 및 특성은 일 실시예에 관련하여 본 발명의 정신 및 범위를 벗어나지 않으면서 다른 실시예로 구현될 수 있다. 또한, 각각의 개시된 실시예 내의 개별 구성요소의 위치 또는 배치는 본 발명의 정신 및 범위를 벗어나지 않으면서 변경될 수 있음이 이해되어야 한다. 따라서, 후술하는 상세한 설명은 한정적인 의미로서 취하려는 것이 아니며, 본 발명의 범위는, 적절하게 설명된다면, 그 청구항들이 주장하는 것과 균등한 모든 범위와 더불어 첨부된 청구항에 의해서만 한정된다. 도면에서 유사한 참조부호는 여러 측면에 걸쳐서 동일하거나 유사한 기능을 지칭한다.

[0034] 이하, 도면들을 참조하여 본 발명의 바람직한 실시예들을 보다 상세하게 설명하기로 한다.

[0035] 도 1은 본 발명의 일 실시예에 따른 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 장치의 블록도이다.

[0036] 본 발명에 따른 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 장치(10, 이하 장치)는 모바일 기기로 촬영된 raw 영상으로부터 DSLR 수준의 화질의 RGB 영상을 얻기 위해 어긋난 데이터 쌍(pair)을 학습하는 학습기이다.

[0037] 도 1을 참조하면, 본 발명에 따른 장치(10)는 제1 U-Net 신경망 모델(200), 제2 U-Net 신경망 모델(400), 손실 함수부(100) 및 영상 학습부(500)를 포함한다. 일 실시예에서, 상기 장치(10)는 영상 변환부(미도시)를 더 포함할 수 있다.

[0038] 상기 제1 U-Net 신경망 모델(200) 및 상기 제2 U-Net 신경망 모델(400)은 각각 다수의 풀링(Pooling) 계층과 다수의 업샘플링(Upsampling) 계층으로 구성되고, 서로 연결되어 W-Net 구조를 형성한다.

[0039] 상기 손실 함수부(100)는 출력 영상과 타겟 영상 사이의 픽셀 손실, 특징 손실 및 색상 손실의 합으로 정의되는 손실 함수를 정의한다.

[0040] 상기 영상 학습부(500)는 상기 손실 함수를 이용하여 상기 제1 U-Net 신경망 모델(200) 및 상기 제2 U-Net 신경망 모델(400)을 통해 서로 다른 카메라로부터 입력 raw 영상과 출력 RGB 영상의 어긋난 데이터 세트를 학습하여, raw 영상과 RGB 영상의 데이터 쌍(pair)으로 정렬한다.

[0041] 일 실시예로서, 상기 입력 raw 영상은 스마트폰으로 촬영된 영상일 수 있고, 상기 출력 RGB 영상은 높은 화질의 DSLR 카메라로 촬영된 영상일 수 있다.

[0042] 상기 영상 변환부는 학습된 raw 영상과 RGB 영상의 데이터 쌍(pair)을 기초로, 실시간으로 raw 영상으로부터 RGB 영상을 출력할 수 있다. 예를 들어, 스마트폰으로 촬영한 raw 영상을 DSLR 카메라로 촬영된 영상 수준으로 출력할 수 있다.

[0043] 본 발명의 상기 장치(10)는 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑을 수행하기 위한 소프트웨어(애플리케이션)가 설치되어 실행될 수 있으며, 상기 제1 U-Net 신경망 모델(200), 상기 제2 U-Net 신경망 모델(400), 상기 손실 함수부(100) 및 상기 영상 학습부(500)의 구성은 상기 장치(10)에서 실행되는 상기 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑을 수행하기 위한 소프트웨어에 의해 제어될 수 있다.

[0044] 상기 장치(10)는 별도의 단말이거나 또는 단말의 일부 모듈일 수 있다. 또한, 상기 제1 U-Net 신경망 모델(200), 상기 제2 U-Net 신경망 모델(400), 상기 손실 함수부(100) 및 상기 영상 학습부(500)의 구성은 통합 모듈로 형성되거나, 하나 이상의 모듈로 이루어 질 수 있다. 그러나, 이와 반대로 각 구성은 별도의 모듈로 이루어질 수도 있다.

[0045] 상기 장치(10)는 이동성을 갖거나 고정될 수 있다. 상기 장치(10)는, 서버(server) 또는 엔진(engine) 형태일

수 있으며, 디바이스(device), 기구(apparatus), 단말(terminal), UE(user equipment), MS(mobile station), 무선기기(wireless device), 휴대기기(handheld device) 등 다른 용어로 불릴 수 있다.

[0046] 상기 장치(10)는 운영체제(Operation System; OS), 즉 시스템을 기반으로 다양한 소프트웨어를 실행하거나 제작할 수 있다. 상기 운영체제는 소프트웨어가 장치의 하드웨어를 사용할 수 있도록 하기 위한 시스템 프로그램으로서, 안드로이드 OS, iOS, 윈도우 모바일 OS, 바다 OS, 심비안 OS, 블랙베리 OS 등 모바일 컴퓨터 운영체제 및 윈도우 계열, 리눅스 계열, 유닉스 계열, MAC, AIX, HP-UX 등 컴퓨터 운영체제를 모두 포함할 수 있다.

[0047] 종래 기술들은 모바일 카메라 영상을 DSLR 카메라 수준으로 화질 개선을 하기 위해 RGB 영상만을 이용하였다. 하지만 RGB 영상은 이미 카메라에 탑재된 영상 신호 처리기(Image signal processor, ISP)를 통하여 raw 센서 데이터를 가공한 것으로, 데이터의 손실이 발생하였을 수 있다.

[0048] 따라서, 본 발명에서는 raw 센서로부터 획득한 영상을 입력으로 사용하여 DSLR 수준의 화질의 RGB 영상을 생성하는 합성곱 신경망 모델을 학습한다.

[0049] 또한, 일반적으로 서로 다른 카메라로 촬영된 영상 쌍(pair)을 정렬시키기 위하여 SIFT 키포인트 매칭을 사용하는데, 이 또한 에러가 존재하기 때문에 완벽하게 정렬되지 않는다.

[0050] 도 2는 어긋난 raw 영상과 RGB 영상의 예시를 보여주는 도면이다. 도 2를 참조하면, 왼쪽은 스마트 폰으로, 오른쪽은 DSLR 카메라로 동일한 씬을 촬영한 영상이나, vertical shift 현상이 나타났다. 이와 같이, 동일한 씬을 서로 다른 카메라로 촬영하였을 때 영상은 완전히 동일할 수 없다.

[0051] 이러한 데이터 셋의 문제를 해결하기 위하여, 본 발명에서는 기존과 다른 신경망 구조와 손실 함수를 사용한다. 이에 대한 자세한 설명은 이하 도면을 참조하여 서술한다.

[0052] 도 3은 본 발명의 사용하는 하나의 신경망 구조를 보여주는 개념도이다. 도 4는 본 발명의 제1 U-Net 신경망 모델 및 제2 U-Net 신경망 모델이 형성하는 두 단계 W-Net 구조를 보여주는 개념도이다.

[0053] 먼저, 본 발명에서 이용하는 신경망 구조는 U-Net 구조이며, 2개의 신경망 모델을 사용한다. 즉, 도 4와 같이 상기 제1 U-Net 신경망 모델(200)과 상기 제2 U-Net 신경망 모델(400)을 연결시킨 신경망 구조를 형성한다.

[0054] 도 3에는 편의상 하나의 신경망 구조만을 도시하였다. 도 3을 참조하면, U-Net은 다수의 풀링(Pooling) 계층과 업샘플링(Upsampling) 계층으로 이루어져 있어 어긋난 데이터로 학습할 때 효과적이다.

[0055] 본 발명은 U-Net 구조를 추가적으로 변형하여 성능을 향상시킨다. 이를 위해, 도 3과 같이 채널 어텐션 메커니즘(channel attention mechanism)을 적용하여 더욱 유용한 특징(features)을 추출하도록 한다(CA-Convs 부분). 또한, 롱 스킵 커넥션(long skip connection)을 추가하여 신경망의 학습을 용이하게 한다.

[0056] 도 4와 같이, 본 발명은 동일하게 개선된 U-Net 2개를 연속적으로 붙여서 두 단계로 구성하고, 이를 W-Net이라 명명한다. 학습 과정은 첫 번째 U-Net(200)을 본 발명에서 제안하는 손실 함수를 이용하여 학습하고, 학습이 완료된 뒤에 두 번째 U-Net(400)을 동일한 손실 함수를 이용하여 학습한다.

[0057] 도 5는 도 1의 손실 함수부의 상세한 블록도이다.

[0058] 도 5를 참조하면, 상기 손실 함수부(100)는 상기 데이터 세트를 출력 영상과 타겟 영상의 픽셀 차이인 픽셀 손실을 계산하는 픽셀 손실부(110), 미리 설정된 레벨 이상의 특징들에 기반한 퍼셉츄얼(perceptual) 손실로부터 특징을 추출하는 특징 손실을 계산하는 특징 손실부(130) 및 RGB 색상 벡터 사이의 코사인 거리로부터 색상 손실을 계산하는 색상 손실부(150)를 포함한다.

[0059] 일 실시예에서, 상기 손실 함수부(100)는 상기 픽셀 손실, 상기 특징 손실 및 상기 색상 손실을 합하여 최종 손실 함수를 정의하는 손실 합성부(170)를 더 포함할 수 있다.

[0060] 이미 설명한 바와 같이, 본 발명에서 제안하는 손실 함수는 3개의 손실의 조합으로 구성된다. 본 발명에서는 신경망 모델의 출력 영상 $\hat{T}$와 타겟 영상 T의 차이를 비교할 때, 데이터의 정렬에 덜 가변적이고 색상 차이에 더 민감한 손실 함수를 사용한다.

[0061] 첫째, 기본적인 픽셀 손실인 L_{pixel} 손실을 사용한다. 픽셀 손실인 L_{pixel}은 아래의 수학식 1과 같이 정의된다.

[0062] [수학식 1]

[0063] $$L_{\text{pixel}} = \left\| \hat{T} - T \right\|_1$$

[0064] 그러나, 데이터의 어긋남 때문에, 픽셀 손실만을 사용하면 흐릿한 결과 영상을 얻게 되므로 아래의 손실 함수를 더 이용한다.

[0065] 둘째, 특징 손실로써 높은 레벨의 특징들에 기반한 퍼셉츄얼 손실을 활용한다. 높은 레벨의 특징들은 여러 번의 다운 샘플링(down sampling)을 거쳐서 생성되기 때문에 데이터의 어긋남에 덜 가변적이다.

[0066] 본 발명에서는 사전에 학습된 VGG-19의 rule4_1 또는 rule5_1 계층에서 특징을 추출할 수 있다. 특징 손실을 픽셀 손실과 함께 사용하면 픽셀 손실만을 사용했을 경우보다 선명한 화질의 결과 영상을 얻을 수 있다. 아래의 수학식 2는 특징 손실을 정의한다.

[0067] [수학식 2]

[0068] $$L_{\text{feat}} = \left\| \emptyset(\hat{T}) - \emptyset(T) \right\|_1$$

[0069] 여기서, $\emptyset$는 VGG-19의 rule4_1 또는 rule5_1 계층을 나타낸다.

[0070] 셋째, raw에서 RGB로의 색상 변환을 효과적으로 학습하기 위해 색상 손실을 사용한다. 색상 손실은 RGB 색상 벡터 사이의 코사인 거리로 정의된다. 색상 손실까지 조합하여 신경망을 학습하면 더욱 정확한 색상의 결과 영상을 얻을 수 있다. 아래의 수학식 3은 색상 손실을 정의한다.

[0071] [수학식 3]

[0072] $$L_{\text{color}} = 1 - \frac{1}{N}\sum_{i}^{N} \frac{\hat{T}_i^{\downarrow 2} \cdot T_i^{\downarrow 2}}{\left\| \hat{T}_i^{\downarrow 2} \right\| \left\| T_i^{\downarrow 2} \right\|}$$

[0073] 여기서, $^{\downarrow 2}$는 영상의 2배 축소를 의미하고, ^{i}는 ^{i}번째 픽셀에 대한 인덱스이다.

[0074] 종합하면, 본 발명에서 제안하는 손실 함수는 상기 픽셀 손실, 상기 특징 손실 및 상기 색상 손실의 합으로 정의되며, 수학식으로 표현하면 아래의 수학식 4와 같다.

[0075] [수학식 4]

[0076] $$L_{\text{total}} = L_{\text{pixel}} + L_{\text{feat}} + L_{\text{color}}$$

[0077] 이와 같이 본 발명에서는 새롭게 도출된 손실 함수를 이용하여 상기 제1 U-Net 신경망 모델(200)을 학습하고, 학습이 완료되면, 상기 제2 U-Net 신경망 모델(400)에서 동일한 손실 함수를 이용하여 학습한다.

[0078] 이러한 학습 결과, raw 영상과 RGB 영상의 데이터 쌍(pair)의 정렬 데이터를 획득할 수 있고, 학습된 raw 영상과 RGB 영상의 데이터 쌍(pair)을 이용하여, 이후 실시간으로 raw 영상으로부터 RGB 영상을 출력할 수 있다.

[0080] 도 6은 본 발명의 일 실시예에 따른 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 방법의 흐름도이다.

[0081] 본 실시예에 따른 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 방법은, 도 1의 장치(10)와 실질적으로 동일한 구성에서 진행될 수 있다.

[0082] 따라서, 도 1의 장치(10)와 동일한 구성요소는 동일한 도면부호를 부여하고, 반복되는 설명은 생략한다. 또한, 본 실시예에 따른 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 방법은 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑을 위한 소프트웨어(애플리케이션)에 의해 실행될 수 있다.

[0083] 본 발명에서는 raw 센서로부터 획득한 영상을 입력으로 사용하여 DSLR 수준의 화질의 RGB 영상을 생성하는 합성곱 신경망 모델을 학습한다.

[0084] 도 6을 참조하면, 본 실시예에 따른 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 방법은, 먼저, 서로 다른 카메라로부터 입력 raw 영상과 출력 RGB 영상의 어긋난 데이터 세트를 획득한다(단계 S10).

[0085] 일 실시예로서, 상기 입력 raw 영상은 스마트폰으로 촬영된 영상일 수 있고, 상기 출력 RGB 영상은 높은 화질의 DSLR 카메라로 촬영된 영상일 수 있다.

[0086] 다수의 풀링(Pooling) 계층과 다수의 업샘플링(Upsampling) 계층으로 구성된 제1 U-Net 신경망 모델을 생성한다(단계 S20).

[0087] 본 발명에서, 제1 U-Net 신경망 모델 및 제2 U-Net 신경망 모델은 각각 다수의 풀링(Pooling) 계층과 다수의 업샘플링(Upsampling) 계층으로 구성되고, 서로 연결되어 W-Net 구조를 형성한다.

[0088] 본 발명은 U-Net 구조를 추가적으로 변형하여 성능을 향상시킨다. 이를 위해, 도 3과 같이 채널 어텐션 메커니즘(channel attention mechanism)을 적용하여 더욱 유용한 특징(features)을 추출하도록 한다(CA-Convs 부분). 또한, 롱 스킵 커넥션(long skip connection)을 추가하여 신경망의 학습을 용이하게 한다.

[0089] 먼저, 출력 영상과 타겟 영상 사이의 픽셀 손실, 특징 손실 및 색상 손실의 합으로 정의되는 손실 함수를 이용하여, 상기 데이터 세트를 상기 제1 U-Net 신경망 모델을 통해 학습한다(단계 S30).

[0090] 본 발명에서 제안하는 손실 함수는 3개의 손실의 조합으로 구성된다. 본 발명에서는 신경망 모델의 출력 영상 $\hat{T}$ 와 타겟 영상 T 의 차이를 비교할 때, 데이터의 정렬에 덜 가변적이고 색상 차이에 더 민감한 손실 함수를 사용한다.

[0091] 첫째, 기본적인 픽셀 손실인 L_{pixel} 손실을 사용한다. 픽셀 손실인 L_{pixel} 은 상기 수학식 1과 같이 정의된다. 그러나, 데이터의 어긋남 때문에, 픽셀 손실만을 사용하면 흐릿한 결과 영상을 얻게 되므로 아래의 손실 함수를 더 이용한다.

[0092] 둘째, 특징 손실로써 높은 레벨의 특징들에 기반한 퍼셉츄얼 손실을 활용한다. 높은 레벨의 특징들은 여러 번의 다운 샘플링(down sampling)을 거쳐서 생성되기 때문에 데이터의 어긋남에 덜 가변적이다.

[0093] 본 발명에서는 사전에 학습된 VGG-19의 rule4_1 또는 rule5_1 계층에서 특징을 추출할 수 있다. 특징 손실을 픽셀 손실과 함께 사용하면 픽셀 손실만을 사용했을 경우보다 선명한 화질의 결과 영상을 얻을 수 있다. 상기 수학식 2는 특징 손실을 정의한다.

[0094] 셋째, raw에서 RGB로의 색상 변환을 효과적으로 학습하기 위해 색상 손실을 사용한다. 색상 손실은 RGB 색상 벡터 사이의 코사인 거리로 정의된다. 색상 손실까지 조합하여 신경망을 학습하면 더욱 정확한 색상의 결과 영상을 얻을 수 있다. 상기 수학식 3은 색상 손실을 정의한다.

[0095] 결론적으로, 본 발명에서 제안하는 손실 함수는 상기 픽셀 손실, 상기 특징 손실 및 상기 색상 손실의 합으로 정의되며, 수학식으로 표현하면 상기 수학식 4와 같다.

[0096] 다수의 풀링 계층과 다수의 업샘플링 계층으로 구성된 제2 U-Net 신경망 모델을 생성한다(단계 S50).

[0097] 새롭게 도출된 손실 함수를 이용하여 상기 제1 U-Net 신경망 모델을 학습하고, 학습이 완료되면, 상기 제2 U-Net 신경망 모델에서 동일한 손실 함수를 이용하여 학습한다.

[0098] 다시 말해, 상기 손실 함수를 이용하여, 상기 제1 U-Net 신경망 모델을 통해 학습된 결과를 상기 제2 U-Net 신경망 모델 학습을 통해 학습하여 raw 영상과 RGB 영상의 데이터 쌍(pair)으로 정렬한다(단계 S60).

[0099] 이에 따라, 학습된 raw 영상과 RGB 영상의 데이터 쌍(pair)을 기초로, 실시간으로 raw 영상으로부터 RGB 영상을 출력할 수 있다. 예를 들어, 스마트폰으로 촬영한 raw 영상을 DSLR 카메라로 촬영된 영상 수준으로 출력할 수 있다.

[0100] 이하에서는, 본 발명의 성능을 평가하기 위한 실험 결과를 기술한다.

[0101] 스마트폰으로 촬영된 입력 raw 영상으로부터 DSLR 수준의 화질을 갖는 RGB 영상을 얻는다. International conference on computer vision(ICCV) 안의 advanced in image manipulation(AIM) raw-to-RGB mapping challenge에서 공개한 Zurich RAW2RGB(ZRR) 데이터셋을 이용하여 본 발명에서 개발한 신경망 모델을 학습하고

테스트를 진행하였다. 성능 평가 지표는 peak signal to noise ratio(PSNR)과 structural similarity index(SSIM)을 사용한다.

[0102] 도 7은 네트워크 구조 향상에 따른 성능 향상을 보여주고, 도 8은 손실 함수의 조합에 따른 성능 향상을 보여준다. 이를 통하여, 본 발명에서 제안한 신경망 구조와 손실 함수가 성능을 많이 향상시켰음을 확인할 수 있다.

[0103] 도 9는 본 발명의 데이터 셋에서의 성능 비교를 보여주는 표이다.

[0104] 도 9를 참조하면, 본 발명이 챌린지에 참여한 다른 기법들에 비하여 높은 성능을 얻은 것을 볼 수 있다. PSNR과 SSIM의 측면에서 가장 높은 성능을 달성하였다. 또 다른 평가 지표인 mean opinion score(MOS)는 사람에 의한 주관 평가로, 1위가 근소한 차이로 2위를 하였다.

[0105] 도 10은 본 발명에 따른 결과의 예시를 보여주는 도면으로, 모바일 기기로 촬영된 raw 영상으로부터 DSLR 수준의 화질의 RGB 영상을 얻는 것을 확인할 수 있다.

[0106] 이와 같은, 두 단계 U-Net 구조 기반의 어긋난 데이터를 이용한 raw에서 RGB로의 매핑 방법은 애플리케이션으로 구현되거나 다양한 컴퓨터 구성요소를 통하여 수행될 수 있는 프로그램 명령어의 형태로 구현되어 컴퓨터 판독 가능한 기록 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능한 기록 매체는 프로그램 명령어, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다.

[0107] 상기 컴퓨터 판독 가능한 기록 매체에 기록되는 프로그램 명령어는 본 발명을 위하여 특별히 설계되고 구성된 것들이거니와 컴퓨터 소프트웨어 분야의 당업자에게 공지되어 사용 가능한 것일 수도 있다.

[0108] 컴퓨터 판독 가능한 기록 매체의 예에는, 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체, CD-ROM, DVD와 같은 광기록 매체, 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 ROM, RAM, 플래시 메모리 등과 같은 프로그램 명령어를 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다.

[0109] 프로그램 명령어의 예에는, 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드도 포함된다. 상기 하드웨어 장치는 본 발명에 따른 처리를 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

[0110] 이상에서는 실시예들을 참조하여 설명하였지만, 해당 기술 분야의 숙련된 당업자는 하기의 특허 청구의 범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위 내에서 본 발명을 다양하게 수정 및 변경시킬 수 있음을 이해할 수 있을 것이다.

산업상 이용가능성

[0112] 본 발명은 어긋난 raw 영상과 RGB 영상의 데이터 쌍(pair)의 정렬 데이터를 획득할 수 있고, 학습된 raw 영상과 RGB 영상의 데이터 쌍(pair)을 이용하여, 이후 실시간으로 촬영한 스마트폰 영상으로부터 DSLR 수준의 영상을 출력할 수 있다.

부호의 설명

[0114] 10: raw에서 RGB로의 매핑 장치

100: 손실 함수부

200: 제1 U-Net 신경망 모델

400: 제2 U-Net 신경망 모델

500: 영상 학습부

110: 픽셀 손실부

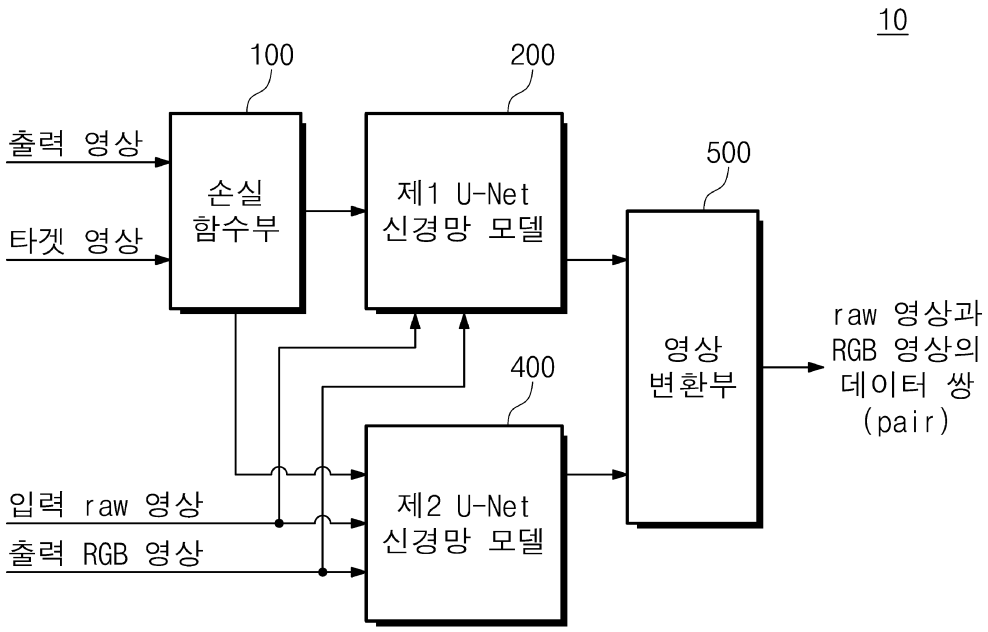
130: 특징 손실부

150: 색상 손실부

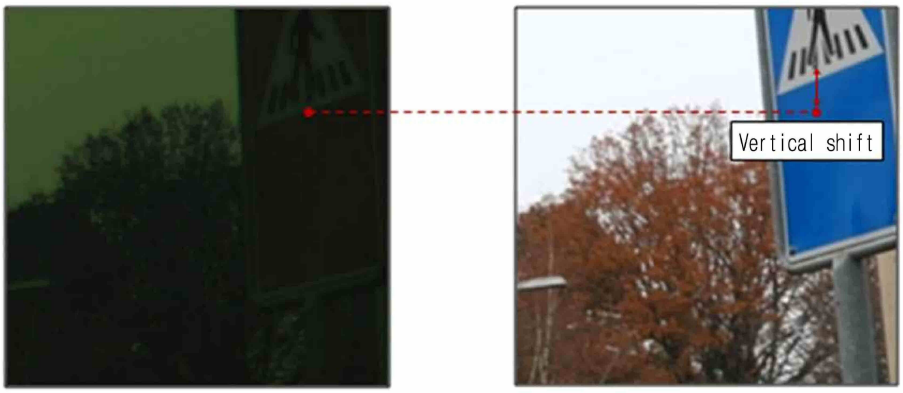
170: 손실 합성부

도면

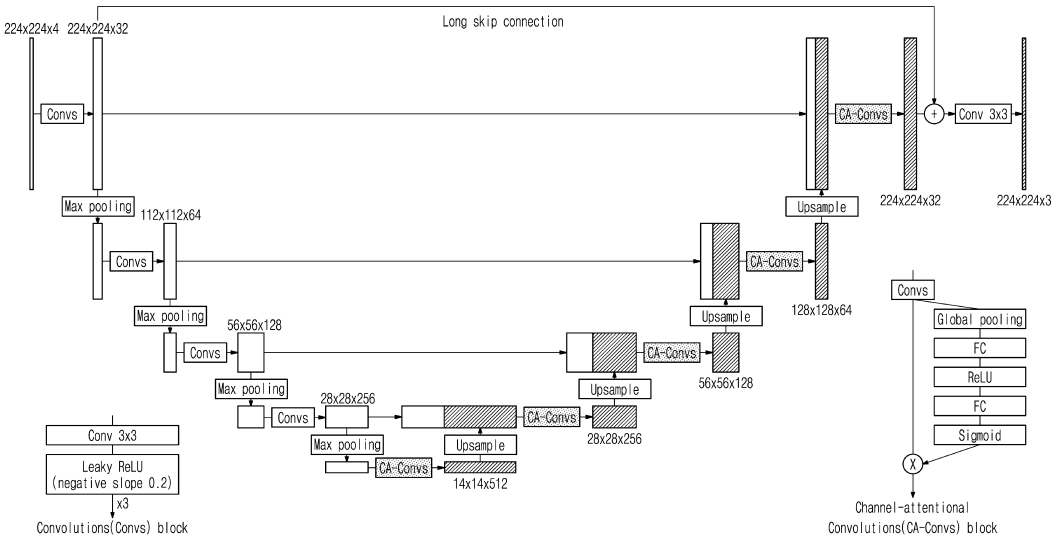
도면1



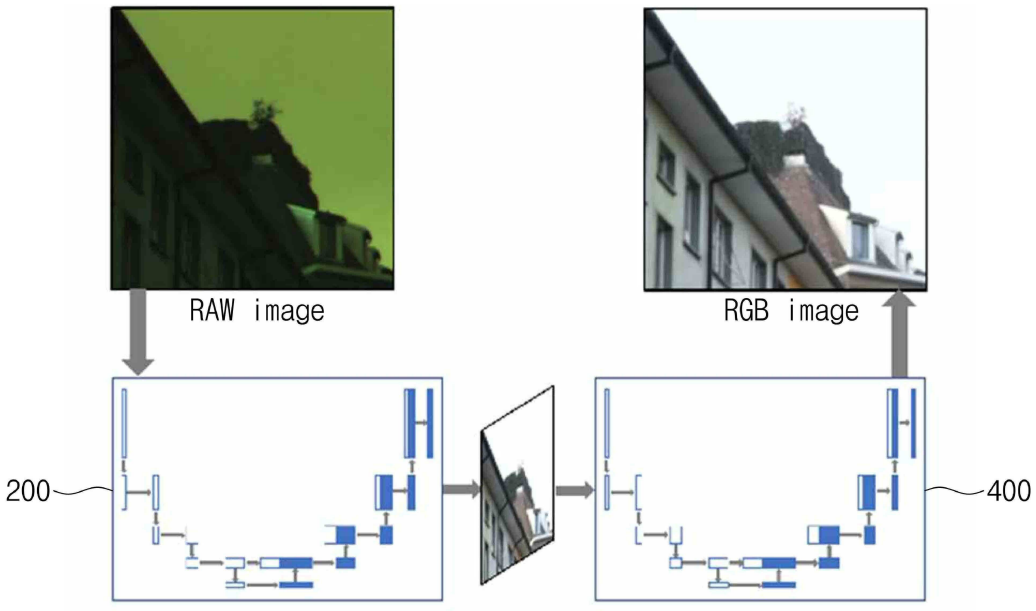
도면2



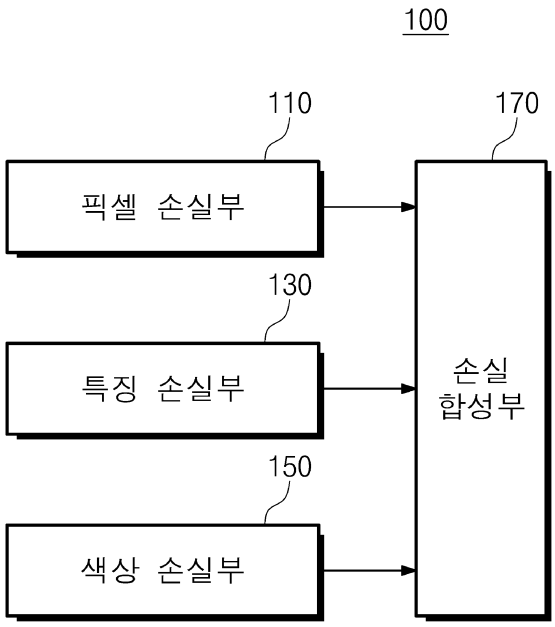
도면3


Long skip connection
224x224x4
224x224x32
Convs
Max pooling
112x112x64
Convs
Max pooling
56x56x128
Convs
Max pooling
28x28x256
Convs
Max pooling
CA-Convs
14x14x512
Upsample
CA-Convs
28x28x256
Upsample
CA-Convs
56x56x128
Upsample
CA-Convs
128x128x64
Upsample
CA-Convs
224x224x32
Conv 3x3
224x224x3
Conv 3x3
Leaky ReLU
(negative slope 0.2)
x3
Convolutions(Convs) block
Convs
Global pooling
FC
ReLU
FC
Sigmoid
Channel-attentional
Convolutions(CA-Convs) block

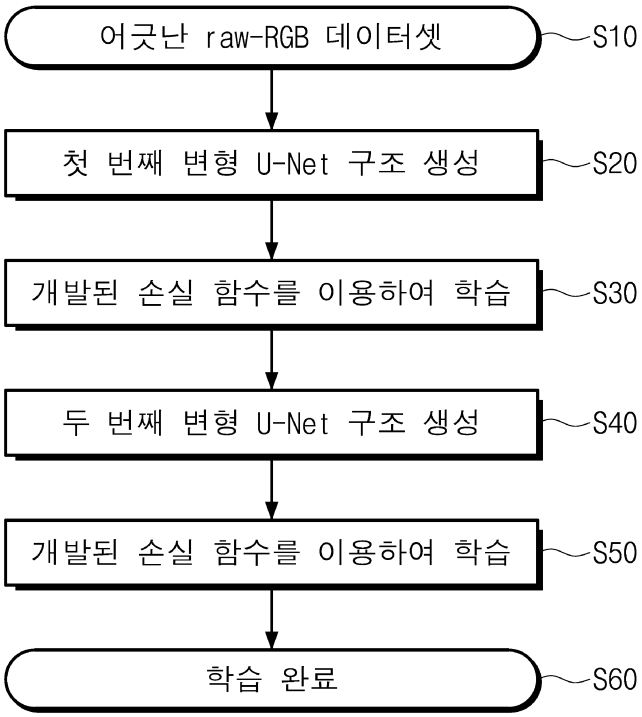
도면4


RAW image
RGB image
200
400

도면5


100
110
픽셀 손실부
130
특징 손실부
150
색상 손실부
170
손실
합성부

도면6


어긋난 raw-RGB 데이터셋
S10
첫 번째 변형 U-Net 구조 생성
S20
개발된 손실 함수를 이용하여 학습
S30
두 번째 변형 U-Net 구조 생성
S40
개발된 손실 함수를 이용하여 학습
S50
학습 완료
S60

도면7

Network enhancements	Baseline: U-Net			
Channel attention module		√	√	√
Long skip connection			√	√
Two-stage structure				√
PSNR	22.30	22.58	22.72	22.82
SSIM	0.8687	0.8704	0.8700	0.8741

도면8

Model	Loss function	VGG layer	PSNR	SSIM
Model 1	L_{pixel}	-	22.82	0.8741
Model 2	$L_{pixel} + L_{feat}$	relu4_1	23.12	0.8755
Model 3	$L_{pixel} + L_{feat}$	relu5_1	23.14	0.8709
Model 4	$L_{pixel} + L_{feat} + L_{color}$	relu4_1	23.18	0.8750
Model 5	$L_{pixel} + L_{feat} + L_{color}$	relu5_1	23.19	0.8719
Ensemble (Model 2 + Model 4 + Model 5)	-	-	23.70	0.8826

도면9

Rank	Track 1: Fidelity			Track 2: Perceptual		Runtime (ms)
	Method	PSNR	SSIM	Method	MOS	
1	W-Net (Ours)	22.59	0.81	1st	1.24	36
2	2nd	22.24	0.80	W-Net (Ours)	1.28	750
3	3rd	21.94	0.79	3rd	1.46	20
4	4th	21.91	0.79	4th	1.56	16
5	5th	19.46	0.75	5th	1.92	20
6	6th	-	-	6th	2.16	8
7	7th	16.86	0.53	7th	-	1280

도면10

