

단노출 플래시 스마트폰 영상에서 저속 동조 영상 생성

이종협[°] 조성현 이승용^{*}

포항공과대학교

{ljh5644, s.cho, leesy}@postech.ac.kr

Slow Sync Image Synthesis from Short Exposure Flash Smartphone Images

Jonghyeop Lee[°] Sunghyun Cho Seungyong Lee^{*}

POSTECH

요약

저속 동조는 촬영자가 장노출과 카메라 플래시를 동시에 이용해서 전경과 배경을 밝게 하는 촬영 기법이다. 단노출 플래시 촬영과 플래시 없는 장노출 촬영과는 달리 저속 동조는 어두운 환경에서의 밝은 전경과 배경을 보장한다. 하지만 스마트폰으로 저속 동조 촬영은 어려운데, 이는 스마트폰 카메라의 플래시는 약한 지속 광이고 노출 시간이 길어지면 플래시를 켜지 못하기 때문이다. 본 연구에서는 단노출 플래시 영상에서 저속 동조 영상을 만드는 딥러닝 방법을 제안한다. 본 연구에서는 공간상에서 가변적인 영상 밝기 개선을 위해 가중치 맵을 적용한 네트워크를 제안한다. 본 연구에서는 지도 학습을 위한 스마트폰 단노출 플래시 영상과 저속 동조 영상 데이터 세트도 제안한다. RAW 영상의 선형성을 이용해 단노출 플래시 영상과 플래시 없는 장노출 영상으로부터 저속 동조 영상을 생성해서 데이터 세트를 구축한다. 실험을 통해 본 연구의 방법이 저속 동조 영상을 효과적으로 생성하는 것을 볼 수 있다.

Abstract

Slow sync is a photography technique where a user takes an image with long exposure and a camera flash to enlighten the foreground and background. Unlike short exposure with flash and long exposure without flash, slow sync guarantees the bright foreground and background in the dim environment. However, taking a slow sync image with a smartphone is difficult because the smartphone camera has continuous and weak flash and can not turn on flash if the exposure time is long. This paper proposes a deep learning method that input is a short exposure flash image and output is a slow sync image. We present a deep learning network with a weight map for spatially varying enlightenment. We also propose a dataset that consists of smartphone short exposure flash images and slow sync images for supervised learning. We utilize the linearity of a RAW image to synthesize a slow sync image from short exposure flash and long exposure no-flash images. Experimental results show that our method trained with our dataset synthesizes slow sync images effectively.

키워드: 저속 동조, 저조도 영상 개선, 딥러닝

Keywords: Slow sync, low-light image enhancement, deep learning

1. 서론

밤과 같이 어두운 환경에서 사진을 촬영하는 것은 흔하지만, 어두운 환경에서 촬영한 사진은 빛의 부족으로 인해 물체 등이 보이지 않는 문제가 발생한다. 빛의 부족을 해결하기 위한 기존의 방식은 ISO 또는 노출 시간을 늘리는 것인데, ISO 증가는 노이즈

증폭을 일으키고, 노출 시간 증가는 카메라 고정을 요구해서 사용자가 불편함을 느낄 수 있다. 또한, ISO나 노출 시간 증가로는 배경에 빛이 있지만, 전경에 빛이 없는 경우는 해결하지 못한다. Fig. 1a에서 노출 시간이 연장되어도 빛의 부족으로 인해 전경의 탁자가 거의 보이지 않는 것을 볼 수 있다. 전경과 배경의 빛의 불균형은 시각적으로 보기 좋은 사진을 찍는 데 있어서 중요한

*corresponding author: Seungyong Lee/POSTECH(leesy@postech.ac.kr)



Figure 1: 어두운 환경에서 촬영된 사진과 밝기 개선 결과물 예시

문제이다.

어두운 환경에서 전경과 배경을 모두 밝게 하는 해결법은 저속 동조 (slow sync)이다. 저속 동조는 장노출과 카메라 플래시를 동시에 사용하는 촬영 기법이다. 저속 동조를 이용하면 카메라 플래시로 전경을 밝히고, 장노출로 배경을 밝힐 수 있다. Fig. 1b와 Fig. 1c는 각각 단노출 플래시 사진과 저속 동조 사진의 예시를 보여준다. 사용자가 단노출과 플래시를 이용해서 밤에 촬영하면, 전경은 밝지만, 배경은 어두운 것을 볼 수 있다. 반면 저속 동조 사진은 전경과 배경이 모두 밝은 것을 볼 수 있다. 하지만, 저속 동조는 전문가 DSLR 카메라와 외부 플래시와 장노출을 요구하기 때문에 스마트폰 사용자는 저속 동조 촬영하기 힘들다.

본 연구의 목표는 스마트폰 단노출 플래시 사진에서 저속 동조 사진을 생성하는 것이다. 입력은 단노출로 촬영되어야 하는데, 장노출로 촬영하는 것은 카메라와 물체의 고정을 요구하기 때문에 사용자가 불편해할 수 있기 때문이다. 단노출 플래시 사진은 어두운 배경과 플래시로 인해 밝은 전경을 지닌다. 이상적인 저속 동조 사진은 전경과 배경이 밝고, 전경의 밝기는 단노출 플래시 사진의 전경의 밝기와 같아야 한다. 따라서, 본 연구의 목표는 스마트폰 단노출 플래시 사진에서 저속 동조 사진을 생성하기 위해 전경의 밝기는 유지하면서 배경의 밝기는 증가시키는 것이다.

본 연구의 목표와 유사한 연구 주제는 저조도 영상 개선 (low-

light image enhancement)이다. 저조도 영상 개선의 목표는 어두운 영상을 밝게 하고 디테일을 향상 시키는 것이다. 이전 저조도 영상 개선 연구는 retinex 이론 [2]를 기반으로 한 최적화를 사용해 이미지에서 조명 (illumination)과 반사율 (reflectance)을 분리하고 조명을 개선해 영상을 밝혔다 [3, 1]. 최근 연구에서는 딥러닝을 이용해 저조도 영상을 개선하는 연구도 제안되었다 [4, 5, 6, 7, 8]. 이전 연구들의 목표는 전체적인 밝기가 낮은 영상을 개선하는 것이 목표이다. 하지만, 단노출 플래시 영상은 플래시로 인해 전경과 배경이 확연히 다른 밝기를 지니고 있다. 그리고 저속 동조 영상 생성은 이미지상의 포화 (saturation)를 방지하기 위해 전경의 밝기는 유지되어야 하지만 Fig. 1d와 같이 이전 연구들은 전경의 밝기 유지를 보장하지 못한다.

본 연구에서는 단노출 플래시 영상에서 저속 동조 영상을 생성하는 딥러닝 방법을 제안한다. 본 연구의 네트워크는 단노출 플래시 영상 전경의 밝기를 유지하면서 배경을 밝게 해서 Fig. 1e와 같은 결과물을 만들어 낸다. 저속 동조 영상 생성은 공간상에서 가변적인 영상 밝기 개선이 필요하므로, 가중치 맵을 도입해 밝기가 증가해야 하는 배경의 어두운 픽셀이 어느 것인지 네트워크에 알려주었다. 본 연구에서는 지도 학습을 위해 스마트폰 단노출 플래시 영상과 저속 동조 영상으로 이루어진 데이터 세트를 수집했다. 스마트폰으로 저속 동조 촬영은 어려운데, 이는 DSLR 플래시와는 달리 스마트폰 카메라는 약한 지속 플래시를 가지고 있고 노출 시간이 길면 플래시를 켜지 못하기 때문이다. 본 연구에서는 RAW 영상의 선형성을 이용해 단노출 플래시 사진과 플래시 없는 장노출 사진으로부터 저속 동조 영상을 생성했다.

본 연구의 기여는 다음과 같다.

- 본 연구에서는 단노출 플래시 영상으로 부터 저속 동조 영상을 생성하는 딥러닝 방법을 제안한다.
- 본 연구에서는 단노출 플래시 영상과 저속 동조 영상으로 구성된 데이터 세트를 수집했다.
- 본 연구는 저속 동조 합성을 위해 최초로 네트워크와 데이터 세트를 제안한다.

2. 관련 연구

사람들이 어두운 환경에서 사진을 촬영하면 사진은 낮은 dynamic range 또는 단노출 또는 낮은 ISO로 인해 어둡게 나올 수 있다. 사진의 어두운 부분에선 내용물이 보이지 않기 때문에 시각적으로 보기 좋지 않다. 저조도 영상 개선의 목표는 저조도 영상의 밝기를 개선하고 디테일을 증가시키는 것이다.

Retinex 이론 [2]을 기반으로 한 최적화 기법들이 저조도 영상 개선을 위해 연구되어왔다 [3, 1]. Retinex 이론 [2]에서는 영상이 물체의 밝기를 나타내는 조명 (illumination)과 물체의 물리적 특성을 나타내는 반사율 (reflectance)의 곱으로 구성된다고 가정한다. SRIE [3]에서는 최적화를 이용해 영상의 조명과 반사율을

예측한 뒤 조명에 감마 보정을 적용해 영상을 밝게 했다. LIME [1]은 적은 계산량을 위해 조명만 예측해서 영상을 밝게 한다.

딥러닝의 발전으로 인해 딥러닝을 이용해 저조도 영상을 개선하는 방법들도 연구되었다 [4, 5, 6, 7, 8]. RetinexNet [8]은 retinex 이론을 기반으로 조명과 반사율을 예측하고 예측된 조명을 개선하는 두 개의 하위 네트워크로 구성되어 있다. Ren 등 [4]은 CNN과 RNN을 활용해서 영상의 디테일을 보존하면서 영상의 밝기를 개선하였다. Wang 등 [6]은 영상의 밝기를 예측하고 이 예측된 밝기의 역과 입력 영상을 곱해서 영상의 밝기를 개선하는 딥러닝 방법을 제안했다. Guo 등 [7]은 딥러닝 모델을 ground truth 없는 손실 함수들만을 이용해서 학습시켰는데, 이들의 모델은 dynamic range 수정을 위한 픽셀 별 곡선 함수를 예측한다. Yifan 등 [5]은 GAN [9]을 이용해 unpaired dataset으로 네트워크를 학습시켰다. 이들의 생성자(generator)는 U-Net [10] 기반이고, 전역 구분자(discriminator)와 지역 구분자를 활용해 지역적으로 다른 밝기 개선을 이루었다.

이전 저조도 영상 개선은 전체적으로 밝기가 낮은 영상들을 개선하였다. 하지만 이전 연구들은 단노출 플래시 영상에서 저속 동조 영상 생성에는 부적합한데, 그 이유는 이전 연구들은 저조도 영상의 모든 픽셀을 개선하기 때문이다. 이전 연구에서 어두운 픽셀들은 밝은 픽셀들보다 더 개선되지만, 밝은 픽셀 또한 개선된다. 저속 동조 영상 생성에서 중요한 점은 배경의 밝기는 올리면서 포화를 방지하기 위해 전경의 밝기는 유지하는 것이다. 하지만, 이전 연구들은 모든 픽셀을 개선하기 때문에 전경 밝기 유지를 보장하지 못한다.

3. 스마트폰 저속 동조 데이터 세트

3.1 스마트폰에서의 저속 동조

본 연구의 목표는 단노출 플래시 사진으로부터 저속 동조 영상을 예측하는 것이다. 그래서 지도 학습을 위해 단노출 플래시 사진과 저속 동조 사진들을 수집해야 한다. 또한, 데이터 세트는 스마트폰으로 수집되어야 한다.

사진작가들은 대부분 DSLR과 외부 플래시와 장노출을 이용해서 저속 동조를 촬영한다. 순간적으로 강력한 빛을 발광하는 DSLR 플래시와는 달리 스마트폰 카메라 플래시는 약하고 지속적인 빛을 발광한다. 또한, 스마트폰 카메라 플래시의 발광 시간은 노출 시간에 비례한다. Fig. 2는 서로 다른 노출 시간에 따른 스마트폰 플래시 사진들을 보여준다. 이 사진들은 매우 어두운 환경에서 Galaxy S21로 촬영되었다. 스마트폰 카메라 애플리케이션은 플래시를 켜고 끄고 할 수 있지만, 플래시 밝기를 조절할 수 없다. 노출 시간이 길어짐에 따라 영상의 밝기가 증가하는데, 이는 스마트폰 플래시가 지속해서 발광하고 이 발광 시간은 노출 시간에 비례하기 때문이다. 노출 시간과 발광 시간 간의 비례 관계로 인해 장노출 촬영 시 카메라 플래시가 전경을 포화를 일으킬 위험이 있다. 또한, 스마트폰 플래시는 노출 시간이 일정 이상일

때 발광하지 못한다. 예를 들어, Galaxy S21의 경우 노출 시간이 1/6초 이상일 때 카메라 애플리케이션은 플래시를 켜지 못한다. 이러한 이유로 인해 스마트폰 카메라 애플리케이션을 가지고 저속 동조 촬영을 하는 것은 어렵다.

본 연구에서는 RAW의 선형성을 이용해 단노출 플래시 영상과 플래시 없는 장노출 영상으로부터 저속 동조 영상을 생성한다. 카메라 센서는 선형적으로 빛의 강도를 측정하고 이 선형 정보를 RAW 영상에 저장한다. 그래서 RAW 영상은 빛의 강도에 선형적이다. 반면 RGB 영상은 빛의 강도에 선형적이지 않다. 영상 신호 처리 파이프라인(image signal process pipeline)은 RAW 영상을 RGB 영상으로 바꾸는데, 영상 신호 처리 파이프라인의 톤 매핑과 감마 보정으로 인해 RGB 영상은 빛의 강도에 선형적이지 않다.

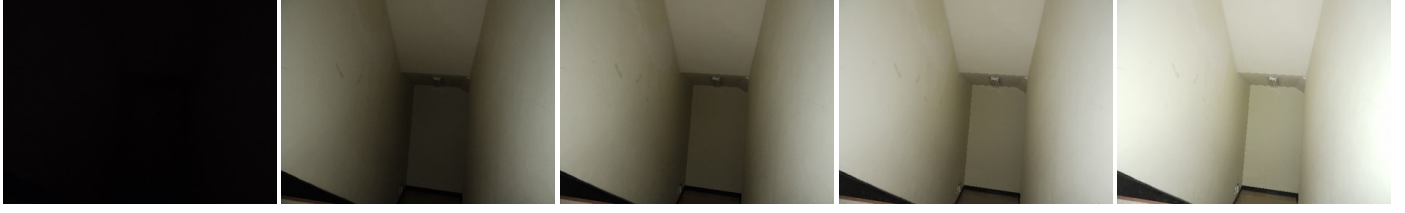
RAW 영상의 선형성으로 인해 RAW 영상에서 특정 빛의 강도를 추출하거나 조정하는 것이 가능하다. Askoy 등 [11]은 플래시 있는 RAW 영상에 플래시 없는 RAW 영상을 빼서 순수 플래시 RAW 영상을 추출해 냈다. 플래시 없는 영상은 배경 빛만 포함하고 있고, 플래시 있는 영상은 배경과 플래시 빛을 포함하고 있으므로, 빼기를 이용해 순수 플래시 영상을 추출할 수 있다.

저속 동조 영상은 단노출의 카메라 플래시 빛과 장노출의 배경 빛을 지녀야 한다. 그래서 본 연구에서는 단노출 순수 플래시 RAW 영상과 플래시 없는 장노출 RAW 영상을 더해서 저속 동조 RAW 영상을 얻는다. 먼저 스마트폰 카메라 애플리케이션을 이용해 단노출 플래시 RAW 영상과 플래시 없는 장노출 RAW 영상을 얻는다. 단노출 순수 플래시 RAW 영상을 얻기 위해 단노출 플래시 RAW 영상에서 선형배율(linear scaling) 적용된 플래시 없는 장노출 RAW 영상을 뺀다. 플래시 없는 장노출 RAW 영상과 단노출 플래시 RAW 영상은 서로 다른 노출 시간의 배경 빛을 지니고 있으므로 선형배율을 통해 두 영상이 같은 배경 빛을 지니게 할 수 있고 빼기를 통해 순수 플래시 영상을 얻을 수 있다. 이러한 RAW를 이용한 방법으로 스마트폰 저속 동조 영상을 생성할 수 있지만, 이 방법은 장노출과 같은 장면에 대한 두 장의 사진을 요구하기 때문에 카메라와 물체를 고정해야 한다. 따라서 네트워크의 데이터 세트 수집을 위해서는 이 방법을 사용하기에는 적합하지만, 일반 촬영자가 사용하기에는 불편함을 느낄 수 있다.

3.2 저속 동조 데이터 세트 구축

단노출 플래시 RAW 영상을 R_{sf} , 플래시 없는 장노출 RAW 영상을 R_l 이라고 가정하자. R_{sf} 과 R_l 은 스마트폰으로 촬영된다. R_l 은 장노출의 배경 빛만 포함하고 있고, R_{sf} 은 단노출의 배경 빛과 플래시 빛을 포함하고 있다. R_{sf} 은 아래와 같이 표현될 수 있다.

$$R_{sf} = R_{se} + R_{sof} \quad (1)$$



(a) 플래시 없음 & 1/6 초 (b) 플래시 있음 & 1/50 초 (c) 플래시 있음 & 1/30 초 (d) 플래시 있음 & 1/15 초 (e) 플래시 있음 & 1/6 초

Figure 2: 플래시 옵션과 노출 시간에 따른 스마트폰 영상들

R_{se} 은 배경 빛만 포함하고 있는 RAW 영상이고, R_{sof} 은 플래시 빛만 포함하고 있는 RAW 영상이다.

R_{se} 와 R_l 은 서로 다른 노출 시간의 배경 빛을 지니고 있다. 선형 배율을 통해 R_l 와 R_{se} 를 같게 만들 수 있다.

$$R_{se} = \frac{T_s}{T_l} R_l \quad (2)$$

T_s 와 T_l 는 각각 단노출 시간과 장노출 시간이다.

R_{sof} 은 아래와 같이 표현될 수 있다.

$$R_{sof} = R_{sf} - R_{se} = R_{sf} - \frac{T_s}{T_l} R_l \quad (3)$$

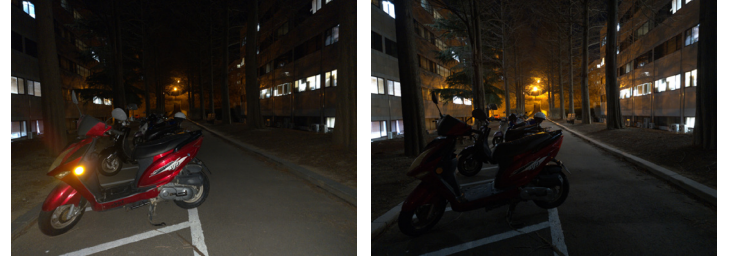
저속 동조 RAW 영상은 단노출 플래시 빛과 장노출 배경 빛을 지니고 있어야 한다. 그래서 저속 동조 RAW 영상은 단노출 순수 플래시 영상과 플래시 없는 장노출 영상의 합이다.

$$R_{slowsync} = R_{sof} + R_l = R_{sf} + (1 - \frac{T_s}{T_l}) R_l \quad (4)$$

$R_{slowsync}$ 은 저속 동조 RAW 영상이다. 영상 신호 처리 파이프라인은 저속 동조 RAW 영상을 저속 동조 RGB 영상으로 변환한다.

Fig. 3는 데이터 세트의 예시를 나타내고 있다. Fig. 3a에서 배경은 어둡지만, 플래시로 인해 전경은 밝다. Fig. 3b에서는 전경은 어둡지만, 장노출로 인해 배경이 밝다. Fig. 3d는 밝은 전경과 배경을 보여준다. Fig. 3c에서 배경이 거의 검은색에 가깝는데 이를 통해 Eq. (3)에서 R_{sof} 를 효과적으로 얻을 수 있음을 알 수 있다. 하지만, R_{sof} 는 카메라 플래시 빛을 제외한 다른 빛을 지니고 있을 수 있다. 예를 들어, Fig. 3c에서 밝은 창문은 제거되지 않았는데 이는 R_{sf} 와 R_l 에서 창문이 포화되었기 때문이다. R_{sof} 가 포화로 인해 제거되지 않은 밝은 빛을 지닐 수 있지만, 이러한 빛은 $R_{slowsync}$ 를 생성하는 데 문제가 되지 않는다. 이는 R_l 이 포화한 픽셀을 지니고 있고, Eq. (4)에서 R_{sof} 의 제거되지 않은 밝은 픽셀과 R_l 의 포화한 픽셀을 더하는 것은 $R_{slowsync}$ 에서도 같은 포화한 픽셀을 만들기 때문이다.

본 연구에서는 Galaxy S21을 사용해서 데이터 세트를 수집했다. Galaxy S21의 카메라 애플리케이션은 노출 시간을 조절할 수 있고 RAW 영상을 저장할 수 있다. 스마트폰을 삼각대에 장착시켜서 방향과 위치를 고정한 뒤, 다양한 노출 시간과 플래시



(a) R_{sf}

(b) R_l



(c) R_{sof}

(d) $R_{slowsync}$

Figure 3: 데이터 세트 예시. 시각화를 위해 영상 신호 처리 파이프라인을 이용해 RAW 영상을 RGB 영상으로 바꾸었다.

설정으로 영상들을 촬영했다. ISO 또는 초점 거리 등은 영상의 품질을 최대화하도록 조정되었다. 단노출 플래시 영상의 노출 시간은 1/45초에서 1/6초 사이이고, 플래시 없는 장노출 영상의 노출 시간은 1/10초에서 2초 사이이다. 장노출 촬영할 경우 물체와 카메라는 고정되어야 하므로 데이터 세트의 물체들은 전부 고정되었다.

본 연구의 데이터 세트는 210장의 플래시 없는 장노출 RAW 영상들과 929장의 단노출 플래시 RAW 영상들로 구성되어 있다. 한 장면에 대해 플래시 없는 장노출 RAW 영상 1장과 단노출 플래시 RAW 영상 여러 장을 찍었는데, 이는 다양한 카메라 노이즈를 수집해서 네트워크를 노이즈에 더 강건해지게 만들기 위함이다. 저속 동조 RGB는 ground truth로, 단노출 플래시 RGB 영상은 입력으로 사용된다. 저속 동조 RGB 영상의 수는 단노출 플래시 RGB 영상의 수와 같다. 20%의 저속 동조 영상들과 단노출 플래시 영상들은 테스트 용도로, 나머지는 훈련 용도로 사용되었다.

4. 저속 동조 생성 네트워크

4.1 네트워크 구조

본 연구의 네트워크는 U-Net [10]을 기반으로 하고 있다. Fig. 4는 본 연구의 네트워크의 구조를 보여주고 있다. U-Net은 영상 처리와 컴퓨터 비전에서 널리 사용되는 네트워크이다. 저조도 영상 개선에서도 U-Net은 활용되는데, 이는 U-Net의 skip connection이 입력 영상의 디테일을 보존하기 때문이다. [5]에 의해 저조도 영상 개선에서 U-Net의 효과가 있음이 보였다. 저속 동조 영상 생성 또한 입력 영상의 디테일 보존이 필요하므로 본 연구에서는 U-Net을 사용했다.

본 연구의 네트워크는 밝기가 증가하여야 하는 픽셀을 강조하는 가중치 맵을 사용한다. 가중치 맵을 만드는 방법은 4.2장에서 설명한다. 본 연구의 네트워크는 단노출 플래시 RGB 영상과 가중치 맵을 입력으로 받고, 저속 동조 RGB 영상을 출력한다. 입력으로 들어간 가중치 맵을 통해 네트워크는 어느 픽셀이 개선되어야 하는지를 알 수 있다. 다운 샘플링된 가중치 맵을 skip connection의 feature map에 곱하는 것도 공간상에서 가변적인 밝기 개선에 도움이 되는데, 이는 가중치 맵의 곱으로 강조된 feature의 픽셀이 개선되어야 한다는 것을 decoder에게 알려주기 때문이다. 단노출 플래시 RGB 영상의 skip connection은 예측된 저속 동조 영상에서 에러를 줄여준다. 이는 ground truth를 바로 예측하는 것보다 ground truth와 입력 영상 간의 차이를 예측하는 것이 더 안정적인 학습에 도움 되기 때문이다. 마지막 합성곱의 결과물과 가중치 맵을 곱하는 것은 전경 밝기의 유지에 도움이 된다. 손실 함수는 ground truth와 예측된 저속 동조 영상 간의 L2 loss이다. 가중치 맵을 통해 네트워크가 저속 동조 영상을 효과적으로 예측할 수 있으므로 추가적인 손실 함수는 필요 없다.

4.2 가중치 맵

가중치 맵은 배경의 어두운 부분은 개선되고 전경의 밝은 부분은 보존되어야 함을 나타낸다. 본 연구에서는 가중치 맵을 이용해 네트워크가 공간상에서 가변적인 밝기 개선을 수행한다. 가중치 맵이 가져야 할 특성들은 다음과 같다.

- 어두운 픽셀들을 강조해야 하므로, 어두운 픽셀의 가중치 값은 커야 하고 밝은 픽셀은 반대로 되어야 한다.
- 가중치 맵을 임의의 feature map과 곱하기 위해 가중치 맵의 채널 차원은 1이 되어야 한다.
- 가중치 맵과 feature map 간의 곱에서 효과적인 강조를 위해 어두운 픽셀과 밝은 픽셀 간의 가중치 값은 충분히 차이가 나야 한다.
- 가중치 맵은 픽셀의 밝기를 나타내야 하기 때문에 공간상에서 부드럽게 변화해야 한다. 노이즈나 텍스처 정보는 가중치 맵에서 제거되어야 한다.

- 가중치 맵의 최솟값은 0에 가까워선 안 된다. 만약 0에 가까우면 feature map과 가중치 맵 사이의 곱으로 인해 feature가 0에 가까워질 수 있다.

본 연구에서는 이러한 특성들을 만족하는 가중치 맵을 만드는 방법을 제안한다. 본 방법의 주요 발상은 입력 단노출 플래시 영상을 회색 영상으로 바꾸어도 플래시의 밝은 전경과 어두운 배경을 보여준다는 점이다. 본 방법은 이 회색 영상을 수정해서 위 특성들을 만족하는 가중치 맵을 만든다.

단노출 플래시 RGB 영상을 I_{sf} 이라 하면 가중치 맵 W 는 다음과 같이 정의된다.

$$W = (1 - RGB2GRAY(filter(I_{sf})))^\alpha * 0.5 + 0.5 \quad (5)$$

$RGB2GRAY$ 는 RGB 영상을 회색 영상으로 바꾸는 함수, $filter$ 는 텍스처 필터 함수, α 는 거듭제곱 매개변수이다. 거듭제곱을 통해 밝은 부분과 어두운 부분의 가중치들을 차이 내서 효과적인 강조를 얻었고, 텍스처 필터 함수를 통해 노이즈와 텍스처를 제거했다. 가중치 맵의 범위는 $[0.5, 1]$ 로 최솟값이 0에 가깝지 않다.

Fig. 5은 α 의 변화에 따른 가중치 맵을 보여준다. 높은 α 의 가중치 맵은 확연한 대비와 공간적 변화를 보여준다. 낮은 α 는 공간적 변화가 거의 없어서, 가중치 맵이 feature를 효과적으로 강조하지 못한다. 본 연구에서는 α 를 4로 설정했는데, 이는 가중치 맵을 시각화 한 결과가 충분한 대비를 보여줬기 때문이다. 더 높은 α 를 사용할 경우 가중치 값들이 최솟값에 가까워져서 정보 손실이 일어날 수 있다.

Fig. 6는 텍스처 필터에 따른 가중치 맵을 보여준다. 텍스처 필터가 없는 가중치 맵은 입력 영상의 텍스처를 가지고 있다. 가중치 맵이 텍스처와 노이즈를 지니고 있으면, 네트워크가 가중치 맵에서 부드러운 밝기 변화가 아닌 텍스처를 보고 잘못 학습할 수 있다. 반면 텍스처 필터를 적용한 가중치 맵은 이러한 텍스처와 노이즈가 없는 것을 볼 수 있다. 본 연구에서는 bilateral filter [12]를 사용했지만, [13]와 같은 다른 텍스처 필터 또한 사용 가능하다.

5. 실험 결과 및 분석

5.1 구현 상세

본 연구에서는 LibRaw 라이브러리 [14]의 Python wrapper인 Rawpy를 사용해서 RAW 영상을 처리했다. 본 연구에서는 네트워크를 28000회 학습했고 batch size는 50, 학습률은 $1e-4$ 로 설정했다. 학습을 위한 입력 영상은 크기가 128에서 384 사이의 무작위 위치의 정사각형으로 잘려진 뒤, 크기가 256이 되도록 재설정되었다. Data augmentation을 위해 영상에 무작위로 뒤집기 또는 회전이 적용되었다. Adam 최적화 [15]를 사용해 훈련했고, 코드는 Pytorch로 구현되었다.



Figure 7: 정성적 결과



Figure 8: 정성적 결과

방법	PSNR	SSIM
SRIE [3]	17.92	0.6253
LIME [1]	11.44	0.3935
Wang 등 [6]	23.41	0.7810
EnlightenGAN [5]	22.58	0.7939
Zero-DCE [7]	8.47	0.2981
Ours	25.08	0.8265

Table 1: 정량적 평가

가중치 입력	가중치 곱	가중치 skip connection	PSNR	SSIM
			24.63	0.8208
✓			24.45	0.8198
	✓		24.61	0.8208
		✓	24.43	0.8213
	✓	✓	24.64	0.8215
✓		✓	24.42	0.8190
✓	✓		24.8	0.8208
✓	✓	✓	25.08	0.8265

Table 2: 네트워크 구조 ablation study의 정량적 결과

5.4 Ablation Study

5.4.1 네트워크 구조

본 연구의 네트워크 구조를 정량적으로 평가해 보았다. 본 네트워크는 가중치 맵과 관련돼서 3가지 요소를 지니고 있다. 첫 번째 요소인 가중치 입력은 네트워크의 입력으로 가중치 맵이 들어가는 것을 의미한다. 두 번째 요소인 가중치 곱은 어두운 배경 강조를 위한 skip connection의 feature map과 가중치 맵 사이의 곱을 의미한다. 마지막 요소인 가중치 skip connection은 전경 밝기 보정을 위한 마지막 합성 곱의 결과물과 가중치 맵 사이의 곱이다.

Table 2은 요소들의 조합들의 정량적 결과를 보여준다. 체크 표시는 해당 요소가 사용되었음을 의미한다. 결과를 통해 더 많은 요소를 사용했을 때 저속 동조 영상을 더 정확히 예측하는 것을 볼 수 있다. 가중치 곱을 홀로 썼을 때는 아무 요소도 사용하지 않을 때와 비슷한 결과가 나오지만 가중치 입력과 가중치 skip connection 요소를 추가하면 성능이 올라간다. 가중치 입력 또는 가중치 skip connection을 홀로 썼을 때는 성능이 낮는데, 이는 충분하지 않은 feature 강조는 오히려 저속 동조 예측에 해가 되는 것으로 예측된다.

5.4.2 가중치 맵

본 연구의 가중치 맵의 효과를 보기 위해 본 연구의 네트워크 또는 가중치를 수정해 4개의 수정된 모델을 실험해 보았다. 첫 번째는 가중치 맵이 없는 네트워크로, 가중치 맵이 없어서 feature map이 강조되지 않는다. 두 번째는 텍스처 필터를 쓰지 않고 낮은 α 를 사용하는 가중치 맵이다. α 를 1로 설정해서 가중치 맵이 작은 공간적 변화량을 지니고 feature map을 효과적으로 강조하지

방법	PSNR	SSIM
가중치 맵 없음	24.63	0.8208
낮은 α 와 텍스처 필터 없음	24.62	0.8199
텍스처 필터 없음	24.7	0.8225
[13] 있음	24.96	0.8225
본 연구 결과	25.08	0.8265

Table 3: 가중치 맵 ablation study의 정량적 평가

못한다. 세 번째는 텍스처 필터를 쓰지 않는 가중치 맵으로, 이 모델을 통해 텍스처 필터의 중요성을 볼 수 있다. 마지막은 interval gradient를 쓰는 텍스처 필터 [13]로 다른 텍스처 필터를 사용해도 좋은 결과가 나오는지 확인하기 위함이다.

Table 3는 가중치 맵 ablation study의 정량적 평가를 보여준다. 낮은 α 또는 텍스처 필터의 부재는 네트워크의 성능을 낮게 만들고, 가중치 맵이 없는 네트워크와 큰 차이가 나지 않게 만든다. 반면 본 연구의 가중치 맵을 사용하면 가장 성능이 잘 나오고 이를 통해 텍스처 필터링과 높은 α 의 중요성을 볼 수 있었다.

6. 결론 및 한계점

본 연구에서는 스마트폰 단노출 플래시 영상에서 저속 동조 영상을 생성하는 딥러닝 방법을 제안한다. 이전 저조도 영상 개선 연구에서는 전경의 포화 및 노이즈 증폭으로 인해 저속 동조 영상을 생성하지 못했다. 본 연구의 네트워크는 배경의 밝기는 올리면서 전경의 밝기를 유지해 저속 동조 영상을 생성할 수 있다. 본 네트워크의 가중치 맵은 공간상에서 가변적인 밝기 개선에도움이 됨을 볼 수 있었다. 본 연구에서는 또한 지도 학습을 위해 단노출 플래시 영상과 저속 동조 영상으로 구성된 데이터 세트를 수집했다. 실험 결과를 통해 본 연구의 방법은 단노출 플래시 영상으로부터 저속 동조 영상을 효과적으로 생성할 수 있음을 볼 수 있었다.

한계점 본 연구의 한계점은 사용자가 배경의 밝기를 조절할 수 없다는 것이다. 단노출 플래시 영상에 대응되는 저속 동조 영상은 다양한 노출 시간에 의해 여러 장이 나올 수 있다. 하지만 본 연구의 방법은 다양한 노출 시간을 고려하지 않고 한 장의 저속 동조 영상만을 만들어 낸다. 밝기를 조절할 수 있는 저속 동조 영상 생성 방법에 관한 연구가 필요함을 알 수 있다.

감사의 글

본 논문은 과학기술정보통신부의 재원으로 정보통신기술진흥센터(SW 스타랩, IITP-2015-0-00174)의 지원을 받아 수행된 연구입니다.

References

- [1] X. Guo, Y. Li, and H. Ling, "Lime: Low-light image enhancement via illumination map estimation," *IEEE Transactions on image processing*, 2016.
- [2] E. H. Land and J. J. McCann, "Lightness and retinex theory," *Josa*, 1971.
- [3] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, "A weighted variational model for simultaneous reflectance and illumination estimation," in *Proc. CVPR*, 2016.
- [4] W. Ren, S. Liu, L. Ma, Q. Xu, X. Xu, X. Cao, J. Du, and M.-H. Yang, "Low-light image enhancement via a deep hybrid network," *IEEE Transactions on Image Processing*, 2019.
- [5] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, and Z. Wang, "Enlightengan: Deep light enhancement without paired supervision," *IEEE Transactions on Image Processing*, 2021.
- [6] R. Wang, Q. Zhang, C.-W. Fu, X. Shen, W.-S. Zheng, and J. Jia, "Underexposed photo enhancement using deep illumination estimation," in *Proc. CVPR*, 2019.
- [7] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong, "Zero-reference deep curve estimation for low-light image enhancement," in *Proc. CVPR*, 2020.
- [8] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," in *BMVC*, 2018.
- [9] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *arXiv*, 2014.
- [10] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*, 2015.
- [11] Y. Aksoy, C. Kim, P. Kellnhofer, S. Paris, M. Elgharib, M. Pollefeys, and W. Matusik, "A dataset of flash and ambient illumination pairs from the crowd," in *Proc. ECCV*, 2018.
- [12] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. ICCV*, 1998.
- [13] H. Lee, J. Jeon, J. Kim, and S. Lee, "Structure-texture decomposition of images with interval gradient," in *Computer graphics forum*, 2017.
- [14] "Libraw," <https://www.libraw.org>, accessed: 2021-04-13.
- [15] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv*, 2014.
- [16] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal processing letters*, 2012.

〈 저 자 소 개 〉



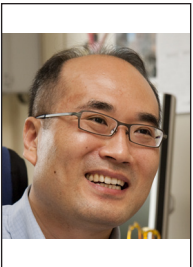
이 중 협

- 2018년 포항공과대학교 컴퓨터공학과 학사
- 2018년~현재 포항공과대학교 컴퓨터공학과 석사과정
- 관심분야: 파노라마, 영상 처리, 딥 러닝 등
- <https://orcid.org/0000-0001-5810-9955>



조 성 현

- 2005년 8월 포항공과대학교 컴퓨터공학과 학사
- 2012년 2월 포항공과대학교 컴퓨터공학과 박사
- 2006년 8월 ~ 2007년 2월 Microsoft Research Asia 인턴
- 2010년 7월 ~ 2010년 11월 Adobe Research 인턴
- 2012년 3월 ~ 2014년 3월 Adobe Research 연구원
- 2014년 4월 ~ 2017년 4월 삼성전자 책임연구원
- 2017년 4월 ~ 2019년 8월 대구경북과학기술원 조교수
- 2019년 8월 ~ 현재 포항공과대학교 조교수
- 관심분야: 영상 및 비디오 처리, 계산 사진학, 영상 복원
- <https://orcid.org/0000-0001-7627-3513>



이 승 용

- 1990년 2월 한국과학기술원 전산학과 석사
- 1995년 2월 한국과학기술원 전산학과 박사
- 1995년 3월~1996년 9월 미국 CUNY 연구원
- 2003년 8월~2004년 7월 독일 MPI fur Informatik 방문 교수
- 2010년 10월~2011년 9월 미국 Adobe Systems 방문교수
- 1996년 10월~현재 포항공과대학교 컴퓨터공학과 교수
- 관심분야: 영상 및 비디오 처리, 비사실적 렌더링, 삼차원 복원
- <https://orcid.org/0000-0002-8159-4271>