

특집논문 (Special Paper)
방송공학회논문지 제24권 제2호, 2019년 3월 (JBE Vol. 24, No. 2, March 2019)
<https://doi.org/10.5909/JBE.2019.24.2.217>
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

효과적인 역 톤 매핑을 위한 필터링 기법

강라훈^{a)}, 박범준^{a)}, 정제창^{a)†}

Image Filtering Method for an Effective Inverse Tone-mapping

Rahoon Kang^{a)}, Bumjun Park^{a)}, and Jechang Jeong^{a)†}

요 약

본 논문에서는 가이드드 영상 필터 (guided image filter: GIF)를 이용하여 컨볼루션 신경망 (convolutional neural network; CNN)을 이용한 역 톤 매핑 (inverse tone-mapping) 기법의 결과를 향상시킬 수 있는 필터링 기법을 제안한다. 저동적범위 (low dynamic range; LDR) 영상을 고동적범위 (high dynamic range; HDR) 디스플레이에서 표현할 수 있도록 변환하는 역 톤 매핑 기법은 지속적으로 제안되어왔다. 최근 들어 컨볼루션 신경망을 이용하여 단일 LDR 영상을 HDR 영상으로 변환하는 알고리즘이 많이 연구되었다. 그 중엔 제한된 동적범위 (dynamic range)로 인해 화소가 포화되어 기존 화소 정보가 손실되는데 이를 학습된 컨볼루션 신경망을 이용해서 복원하는 알고리즘이 존재한다. 해당 알고리즘은 비포화 영역의 잡음까지는 억제하지 못하며 포화 영역의 디테일까지는 복원하지 못한다. 제안한 알고리즘은 입력 영상에 가중된 가이드드 영상 필터 (weighted guided image filter; WGIF)를 사용해서 비포화 영역의 잡음을 억제하고 포화 영역의 디테일을 복원시킨 다음 컨볼루션 신경망에 인가하여 최종 결과 영상의 품질을 개선하였다. 제안하는 알고리즘은 HDR 정량적 화질평가 지표를 측정하였을 때 기존의 알고리즘에 비해 높은 화질평가 지수를 나타내었다.

Abstract

In this paper, we propose a filtering method that can improve the results of inverse tone-mapping using guided image filter. Inverse tone-mapping techniques have been proposed that convert LDR images to HDR. Recently, many algorithms have been studied to convert single LDR images into HDR images using CNN. Among them, there exists an algorithm for restoring pixel information using CNN which learned to restore saturated region. The algorithm does not suppress the noise in the non-saturation region and cannot restore the detail in the saturated region. The proposed algorithm suppresses the noise in the non-saturated region and restores the detail of the saturated region using a WGIF in the input image, and then applies it to the CNN to improve the quality of the final image. The proposed algorithm shows a higher quantitative image quality index than the existing algorithms when the HDR quantitative image quality index was measured.

Keyword : Inverse Tone-mapping, HDR, CNN, Deep Learning, Guided Image Filter

a) 한양대학교 전자컴퓨터통신공학과(Department of Electronics and Computer Engineering, Hanyang University)

† Corresponding Author : 정제창(Jechang Jeong)

E-mail: jjeong@hanyang.ac.kr

Tel: +82-2-2220-4372

ORCID: <https://orcid.org/0000-0002-3759-3116>

※ 본 연구는 미래창조과학부 및 정보통신기술진흥센터의 정보통신·방송 연구개발사업의 일환으로 수행하였음.[2014-0-00670, ICT 장비용 SW 플랫폼 구축]

※ This work was supported by the ICT R&D program of MSIP/IITP.[2014-0-00670, Software Platform for ICT Equipment]

※ 이 논문의 연구결과 중 일부는 “2018년 한국방송·미디어공학회 추계학술대회”에서 발표한 바 있음.

· Manuscript received January 8, 2019; Revised March 18, 2019; Accepted March 18, 2019.

Copyright © 2019 Korean Institute of Broadcast and Media Engineers. All rights reserved.

“This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered.”

I. 서론

최근 들어 디스플레이 산업이 발전하면서 과거에 비해 다양한 기능을 탑재한 디스플레이가 많이 출시되고 있다. 디스플레이 장치는 해상도가 높을 뿐만 아니라 넓은 동적영역을 가진 고동적범위 (high dynamic range; HDR) 영상을 출력할 수 있는 기능을 포함하고 있다. HDR 영상 콘텐츠를 체감할 수 있는 디바이스는 증가하고 있지만 기존의 카메라가 아닌 특수하게 고안된 HDR 카메라를 통해 콘텐츠를 얻기 때문에 이용할 수 있는 콘텐츠의 수는 디바이스에 비해 매우 부족하다. 이를 해결할 방법으로 다양한 노출값을 가진 저동적범위 (low dynamic range ;LDR) 영상을 합성하여 넓은 동적범위로 확장시키는 다중 노출 융합 (multi exposure fusion; MEF) 기법이 제안되었다^[7]. 이 방법은 낮은 연산량을 나타내기 때문에 스마트폰 카메라 등 연산 성능이 낮은 휴대용 디스플레이에 가장 널리 보급된 HDR 영상 기법이다. 하지만 각기 다른 노출값을 가진 영상을 얻기 위해 연속으로 영상을 얻어내는 과정에서 시간 흐름에 따라 움직이는 피사체가 투영되는 고스트 아티팩트 (ghost artifact)가 발생한다^[21]. 또한 미리 설정한 노출값 외에는 다른 노출에서 관측할 수 있는 영역은 포함하지 못한다. 따라서 LDR 영상과 HDR 영상간의 변환 연관성을 이용하여 LDR 영상을 HDR로 변환시키는 방법이 제안되었다. HDR 영상을 LDR로 매핑 시키는 기법을 역 톤 매핑 (inverse tone-mapping)이라고 한다.

기존의 역 톤 매핑 알고리즘은 다음과 같다. 대부분 카메라 곡선을 이용하여 LDR 영상을 비선형적으로 매핑하여 HDR 영상을 획득한다. 기존의 역 톤 매핑 알고리즘들은 어두운 영역을 자세하게 표현하도록 매핑한다. 그로 인해 영상 내 밝은 영역에서 임계치를 넘어간 화소들이 포화된다. LDR 영상에서 한정된 동적범위로 인해 포화돼버린 화소 정보는 매핑 과정을 통해 복원되지 않는다. 기존의 알고리즘을 이용하여 포화되는 영역의 손실된 화소를 원본 화소로 복원할 수 없기 때문에 딥 러닝 (deep learning)을 기반으로 하는 복원 알고리즘이 제안되었다^[5,6]. 그럼에도 불구하고 딥 러닝 기반의 복원 알고리즘도 매핑하는 과정에서 에지 영역의 디테일이 손실되는 문제가 발생하였다. 본 논문에서는 이러한 문제를 해결하기 위해 가중된 가이드드

영상 필터 (weighted guided image filter; WGIF)를 전처리 과정에 적용하였다. WGIF는 에지 영역을 보존하면서 잡음도 억제할 수 있기 때문에 본 논문에서 포화 영역을 제대로 복원하기 위해 사용하였다. 역 톤 매핑 신경망은 포화되는 영역의 손실된 화소를 복원하기 위해 WGIF 결과 영상에 적용되었다.

본 논문은 다음과 같이 구성된다. 2장에서는 관련 이론에 대하여 설명하고 3장에서는 제안하는 알고리즘을 자세히 설명한다. 4장에서는 기존의 알고리즘들과 제안된 알고리즘을 비교하여 성능을 평가하고 5장에서 결론을 맺는다.

II. 관련 이론

1. 역 톤 매핑 (inverse tone-mapping)

기존의 역 톤 매핑 알고리즘은 전역적 방법과 지역적 방법으로 나뉜다. 전역적 방법들은 모든 화소 동적범위를 동일하게 확장시키는 방법이다. Landis 등에 의해 제안된 알고리즘은 멱함수 (power function)를 이용하여 역 톤 매핑을 하였다^[1]. Akyüz 등에 의해 제안된 방법은 감마 보정이 결합된 선형 변환을 사용하였다^[2].

지역적 방법은 주로 분석함수를 통해 확장 맵을 추정하여 역 톤 매핑을 하는 방법들이다. Banterle 등이 제안한 알고리즘은 톤 매핑의 역 과정을 적용하는 밝기를 매핑하고, 확장 맵을 측정하여 광원의 밀도를 추정하였다^[3]. 이 알고리즘은 톤 매핑의 역 과정이기 때문에 역 과정을 적용할 톤 매핑 연산자 (tone-mapping operator; TMO)의 영향을 받는다. 또한 기본적으로 해당 영상이 어두운 영역과 밝은 영역이 균형 잡힌 영상이라는 가정이 되어있는 경우에만 적용된다. 따라서 화소가 포화될 경우 해당 화소를 복원할 수 없다. 지역적 방법 중에서 확장 맵을 추정할 때 대비를 유지하기 위한 에지-스탑핑 (edge-stopping) 함수와 결합된 가우시안 필터를 사용하기도 한다^[4,5,6].

최근 들어 딥 러닝을 이용한 역 톤 매핑 알고리즘이 제안되었다. Eilersten 등이 제안한 알고리즘은 기존의 역 톤 매핑 기법들이 복원하지 못했던 포화 영역 복원에 중점을 두었으며 포화된 영역을 합성곱 신경망 (convolutional neural

network; CNN)으로 검출한 뒤 이를 학습하여 복원한다^[5]. Marnierides 등에 의해 제안된 ExpandNet은 LDR 영역의 지역적 특징을 추출하는 네트워크, 전역적 특징을 추출하는 네트워크 및 HDR 영역으로 팽창시키는 네트워크를 각각 CNN으로 구성하고 그 출력을 하나로 연결하여 HDR 영상으로 복원한다^[6].

2. 가이드드 영상 필터 (guided image filter; GIF)

GIF는 지역적 필터링 기반 에지 보존형 스무딩 기법이다^[20]. GIF는 양방향 필터에 비해 연산량이 낮고 우수한 에지 보존 성능을 나타내지만 후광 아티팩트 (halo artifact)가 발생한다^[20]. 이 문제를 해결하기 위해 Li 등에 의해 WGIF가 제안되었다^[19]. WGIF의 복잡도는 GIF와 같으나 GIF에서 발생하는 후광 아티팩트를 억제할 수 있다. 또한 WGIF는 잡음을 억제할 뿐 아니라 디테일을 개선할 수도 있다. WGIF에서는 에지 영역을 보존하기 위해 가중치를 사용한다(식 1). G 가 가이드드 영상이고 $\sigma_{G,1}^2(p')$ 를 3×3 윈도우의 가이드드 영상의 분산이라고 하자.

$$\Gamma_G(p') = \frac{1}{N} \sum_{p=1}^N \frac{\sigma_{G,1}^2(p') + \epsilon}{\sigma_{G,1}^2(p) + \epsilon} \quad (1)$$

여기서 ϵ 은 입력 영상의 동적범위 L 로부터 계산되어 $(0.001 \times L)^2$ 의 값을 가지는 상수이다. 식 1에서 $\Gamma_G(p')$ 값은 p' 이 에지 영역에 있을 경우 1보다 크게 되고, 부드러운 영역에 있을 경우 1보다 작게 된다. 따라서 에지 영역일 경우 다른 화소들에 비해 가중치를 크게 가지게 된다.

WGIF는 필터링된 결과 영상과 가이드드 영상 간에는 지역적 선형 모델을 만족한다고 가정한다. 따라서 필터링

된 결과 영상의 에지 영역은 가이드드 영상의 에지 영역에 대응된다. 에지 보존 가중치는 비용 함수 $E(a_{p'}, b_{p'})$ 로 통합된다. 결과값은 필터링 결과 영상 \hat{Z} 와 입력 영상 X 사이의 차이값을 최소화함으로 얻어진다. 비용 함수 E 는 다음과 같이 정의된다.

$$E = \sum_{p \in \Omega_{\zeta_1}(p')} [a_{p'} G(p) + b_{p'} - X(p)]^2 + \frac{\lambda}{\Gamma_G(p')} a_{p'}^2 \quad (2)$$

$\Omega_{\zeta_1}(p')$ 는 화소 p' 를 중심으로 반경 ζ_1 을 가진 사각 윈도우를 나타낸다. $a_{p'}$ 와 $b_{p'}$ 는 $\Omega_{\zeta_1}(p')$ 안의 두 상수이다. 최적값 $a_{p'}$ 와 $b_{p'}$ 는 식 3, 4로 표현된다.

$$a_{p'} = \frac{\mu_{G \odot X, \zeta_1}(p') - \mu_{G, \zeta_1}(p') \mu_{X, \zeta_1}(p')}{\sigma_{G, \zeta_1}^2(p') + \frac{\lambda}{\Gamma_G(p')}} \quad (3)$$

$$b_{p'} = \mu_{X, \zeta_1}(p') - a_{p'} \mu_{G, \zeta_1}(p') \quad (4)$$

\odot 는 두 행렬의 요소별 곱셈을 말하고 μ 는 각 항목의 평균값을 의미한다. 최종 필터링된 영상 $\hat{Z}(p)$ 은 다음과 같이 얻어진다.

$$\hat{Z}(p) = \bar{a}_p G(p) + \bar{b}_p \quad (5)$$

\bar{a}_p 와 \bar{b}_p 는 각각 윈도우 안에서 $a_{p'}$ 와 $b_{p'}$ 의 평균값이다. 간단한 계산을 위해 가이드드 영상 G 와 필터링된 영상이 같다고 가정하고 계산하여 사용한다. 입력 영상은 필터링된 베이스 레이어와 디테일 레이어로 이뤄진다. 입력 영상

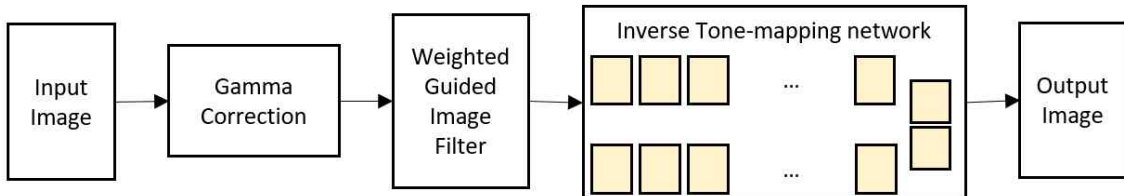


그림 1. 제안하는 알고리즘의 순서도
 Fig. 1. The flow chart of proposed algorithm

에 디테일 레이어의 배수를 곱해 더해주면 디테일이 개선된 영상을 얻을 수 있다 (식 6).

$$Z_{enh}(p) = X(p) + \theta e(p) \tag{6}$$

$Z_{enh}(p)$ 는 디테일이 개선된 영상을 의미하며, $X(p)$ 는 입력 영상, $e(p)$ 는 디테일 레이어를 나타낸다. $e(p)$ 는 입력 영상에서 필터링된 영상 $\hat{Z}(p)$ 를 빼서 얻을 수 있다. θ 는 증폭 인자로 양수 상수이다. 즉, θ 값에 따라 디테일 개선 정도를 조절할 수 있다.

III. 제안하는 알고리즘

제안하는 알고리즘은 새롭게 제안하는 전처리 필터와 역 톤 매핑 컨볼루션 신경망을 결합한 형태이다 (그림 1). 기존의 많은 역 톤 매핑 알고리즘들은 비선형 보정을 고려하지 않아 카메라가 획득한 정확한 화소값을 추정하지 못했다. 따라서 역 감마 보정을 이용하여 선형적으로 재매핑하여 이를 해결할 수 있다. 또한 포화된 영역을 매핑하는 과정에서 디테일이 사라지고 포화된 화소로 인한 비포화 영역에 잡음이 발생한다. 이를 해결하기 위해 에지와 디테일을 보존하고 잡음을 억제하는 WGIF를 전처리 과정에 사용하였다. 전처리 필터에서 입력 LDR 영상이 역 감마 보정을 하고 WGIF를 거친 뒤 최종적으로 전처리 필터를 통과한 결과 영상을 역 톤 매핑 컨볼루션 신경망에 인가하여 복원된 HDR 영상을 출력한다.

1. 전처리 필터

HDR 영상을 선형적으로 톤 매핑하게 되면 우리가 실제로 보는 영상과는 다른 영상을 얻게 된다. 인간의 시각은 베버의 법칙(Weber's law)에 따라 밝기에 대해 비선형적으로 반응하기 때문이다^[8]. 따라서 일반적으로 선형적으로 얻어지는 카메라 영상을 인간이 보는 디스플레이에 출력하기 위해 비선형적으로 매핑하는 과정을 거친다. 보통 어두운 부분을 더 자세하게 표현하기 위해 감마 보정 (gamma correction)을 많이 사용한다. 이를 HDR 영상에서 사용하기 위해서는 역 감마 보정을 하여야 한다. $\hat{D}_{i,c}$ 가 감마 보정된 입력 LDR 영상 화소이고 HDR 영상 화소는 $D_{i,c}$ 일 때 이를 선형화하기 위해 역 감마 보정을 한다 (식 7). 여기서 $\gamma=2.0$ 이다.

$$D_{i,c} = \hat{D}_{i,c}^\gamma \tag{7}$$

그리고 역 감마 보정을 거쳐 비선형적으로 매핑한 영상을 WGIF에 통과시킨다. 제안하는 알고리즘은 WGIF의 디테일을 개선 기능을 이용하여 에지를 보존하고 잡음을 억제할 뿐 아니라 디테일 영역을 개선한다 (식 6).

2. 역 톤 매핑 신경망

역 톤 매핑 신경망은 Eilersten 등이 제안한 구조를 사용하였다. 해당 구조는 포화된 LDR 영상의 영역을 복원하는데 중점을 두었다. 컨볼루션 신경망은 포화된 LDR 영역과

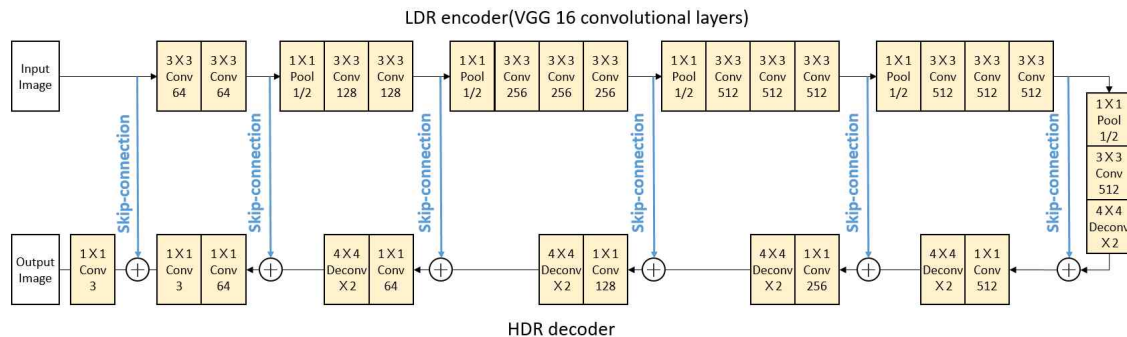


그림 2. 컨볼루션 신경망을 사용한 역 톤 매핑 구조
 Fig 2. The structure of inverse tone-mapping using convolutional neural network

원본 HDR 영역간의 연관성을 학습한다. 이 신경망을 복합 동적범위 오토인코더 (hybrid dynamic range autoencoder) 라고 명명하였다. Hinton 등이 제안한 오토인코더 구조와 비슷하게 인코더, 디코더 구조로 나누어져있다^[9] (그림 2). 오토인코더 구조의 특징은 입력 영상을 저해상도로 부호화하면서 각 해상도에서 특징을 추출한다. 그리고 디코더에서 다시 원래 해상도로 복원시킨다. 이러한 구조를 사용하면 원본 영상의 해상도에서 발견할 수 없었던 각기 다른 해상도에서의 특징을 추출해낼 수 있고, 입력영상과 출력영상의 해상도가 동일하게 출력된다는 장점이 있다. Vincent 등이 제안한 잡음제거 오토인코더는 잡음이 섞인 입력 영상들을 학습하여 원본의 깨끗한 영상을 복원하였다^[10]. Eilersten 등은 해당 잡음 제거 오토인코더 구조에서 포화된 하이라이트 영역을 잡음으로 여기고, 인코더는 LDR 영상으로부터 확장된 하이라이트 부분을 재현하고, 디코더는 이를 HDR로 확장시키는 역할을 하게끔 제안되었다. Eilersten 등은 해당 잡음 제거 오토인코더에서 포화되어버린 하이라이트 영역을 잡음으로 여기고, 인코더는 LDR 영상으로부터 포화 영역을 복원하고, 디코더는 이를 HDR로 확장시키는 역할을 하게끔 제안하였다. 인코더와 디코더 간에는 skip-connection을 이용하여 HDR 복원 과정에서 손실될 수 있는 바로 직전 레이어의 디테일을 인코더 부분에서 바로 전달받는다. 이러한 skip-connection은 He 등이 제안한 네트워크에서 디테일 복원 성능을 올리는 것으로 입증된 기법이다^[11].

인코더는 전결합층 (fully connected layer)을 제외한 VGG16 신경망 구조를 사용하였다^[12]. VGG16 신경망은 3×3 컨볼루션만 사용하여 간단하고 복잡도가 낮다. 디코더는 LDR 영상을 HDR 영상으로 확장하는 부분이기 때문에 log 도메인에서 동작한다. 이는 원본 HDR 영상과 예측된 HDR 영상 간의 손실 함수를 구할 때 원본 HDR 영상의 log 값을 사용하기 때문이다. 디코더에서 출력 영상을 업샘플링 (upsampling) 하기 위해 디컨볼루션 (deconvolution) 을 사용하는데 양선형 보간법 (bilinear interpolation)을 사용하였다. 네트워크의 모든 층들은 rectified linear unit (ReLU) 활성화 함수를 사용하였고^[17], 디코더 각 레이어 뒤

에는 batch normalization을 사용하였다. 그리고 학습과정에서 사용된 손실 함수는 식 8과 같이 정의하였다.

\hat{y} 와 H 는 각각 예측된 HDR 영상과 실제 원본 HDR 영상이다. I 와 R 은 각각 연산된 조명 성분과 반사 성분을 의미한다. y 는 $y = \log(H + \epsilon)$ 식을 만족하며 log 도메인에서 각각 원본과 예측된 값 간의 오차 ϵ 를 구한다. λ 는 조명 성분의 가중치를 의미한다. HDR 영상은 이 신경망을 거쳐서 예측된 값을 선형화된 입력 영상과 결합하여 얻을 수 있다. 결합식은 다음과 같이 나타낼 수 있다 (식 9).

$$\hat{H}_{i,c} = (1 - \alpha_i) f^{-1}(D_{i,c}) + \alpha_i \exp(\hat{y}_{i,c}) \quad (9)$$

$\hat{H}_{i,c}$ 는 복원된 최종 HDR 영상의 컬러 채널 c 의 i 번째 화소 값을 의미한다. $D_{i,c}$ 는 입력 LDR 영상 화소값이고 $\hat{y}_{i,c}$ 는 인코더에서 예측된 화소값이다. f^{-1} 은 역 카메라 곡선이며 실험에선 시그모이드 (sigmoid) 함수를 사용하였다. α_i 는 복원된 영역만을 알파 혼합하기 위한 혼합 계수로 임계치를 넘어선 포화 영역을 정의한다. α_i 를 구하는 식은 다음과 같다 (식 10).

$$\alpha_i = \frac{\max(0, \max(D_{i,c}) - \tau)}{1 - \tau} \quad (10)$$

여기서 τ 는 임계치 값으로 0과 1사이의 값으로 나타나고 히스토그램의 포화 임계치를 의미한다. 예를 들어 0.9로 설정하면 영상 히스토그램의 상위 10% 이상이 포화되었다는 것을 의미한다.

IV. 실험 결과

컨볼루션 신경망을 학습시키기 위해서는 HDR 영상과 이에 상응하는 LDR 영상이 존재해야한다. 기존의 HDR 영상들은 짝지어진 LDR 영상이 대부분 존재하지 않는다. 그래서 OpenCV를 이용하여 HDR영상의 무작위 영역을 무작위 카메라 보정을 적용하여 다양한 훈련 데이터를 얻었다.

$$\mathcal{L}_{LR}(\hat{y}, H) = \frac{\lambda}{N} \sum_i \left| \alpha_i (\log(I_i^{\hat{y}}) - \log(I_i^y)) \right|^2 + \frac{1 - \lambda}{3N} \sum_{i,c} \left| \alpha_i (\log(R_{i,c}^{\hat{y}}) - \log(R_{i,c}^y)) \right|^2 \quad (8)$$

카메라 보정 조건은 각기 다른 화소 포화 비율, 가우시안 잡음, 명도, 채도로 데이터를 증강시키고 무작위 시그모이드 함수를 이용하여 LDR로 톤 매핑하였다. 컨볼루션 신경망 학습용 데이터는 Fairchild 데이터셋과 Ward 데이터셋을 사용하였다^[13,14]. 각각의 데이터셋에서 320×320의 단위 크기로 무작위 위치에서 추출, 회전, 반전시킨 뒤 무작위 시그모이드 함수를 이용해 톤 매핑을 하여 훈련 데이터를 증강시켰다. 그리고 시험에는 훈련에 사용되지 않은 28개의 HDR 영상과 5%만 포화시켜 LDR로 톤 매핑한 영상을 사용하였다. 즉 $\tau = 0.95$ 로 설정하였다.

표 1. 시험 데이터셋에 대한 지각적 연속성 부호화 최대 신호 대 잡음비
Table 1. The PU-PSNR of test dataset

No.	Akyüz et al.	Huo et al.	Eilersten et al.	Expand-Net	Proposed algorithm
1	7.695	7.783	32.341	28.381	32.352
2	10.328	12.639	33.027	30.205	33.227
3	8.771	9.782	33.127	30.987	33.192
4	7.238	6.730	31.265	27.888	31.113
5	8.097	8.317	33.855	23.942	33.823
6	6.973	6.183	28.535	24.745	28.569
7	8.542	9.960	32.773	27.873	32.727
8	11.395	14.836	40.795	37.788	40.662
9	9.698	11.005	35.014	38.394	35.078
10	8.882	9.760	35.330	39.514	35.333
11	9.520	10.617	37.899	38.085	37.900
12	8.444	9.551	36.211	28.776	36.191
13	7.381	7.110	30.564	36.996	30.569
14	8.912	10.558	30.186	30.254	30.170
15	7.398	7.088	25.828	26.366	25.868
16	9.044	10.112	36.757	37.056	36.854
17	8.658	9.648	31.982	33.272	31.984
18	8.112	8.518	35.991	35.970	35.997
19	8.398	9.200	36.472	35.643	36.424
20	7.466	7.164	27.995	38.269	28.015
21	8.250	8.535	36.686	37.626	36.699
22	8.402	9.051	34.734	38.707	34.761
23	9.682	11.032	41.989	39.646	41.840
24	8.294	8.897	32.166	33.744	32.235
25	7.402	7.126	32.550	26.980	32.568
26	7.291	7.059	26.801	25.034	26.842
27	7.480	7.262	32.596	39.307	32.606
28	7.408	6.708	35.390	26.389	35.350
avg.	8.399	9.008	33.531	32.780	33.534

표 2. 시험 데이터셋에 대한 지각적 연속성 부호화 구조적 유사성
Table 2. The PU-SSIM of test dataset

No.	Akyüz et al.	Huo et al.	Eilersten et al.	Expand-Net	Proposed algorithm
1	-0.1714	-0.1541	0.7731	0.8265	0.7738
2	-0.0910	-0.1107	0.9602	0.8974	0.9593
3	-0.1299	-0.1268	0.8804	0.7894	0.8804
4	-0.0917	-0.1151	0.7471	0.6367	0.7480
5	-0.1297	-0.1232	0.7865	0.8104	0.7890
6	-0.1288	-0.1467	0.5357	0.8423	0.5403
7	-0.1377	-0.1428	0.9094	0.8152	0.9095
8	-0.2057	0.0119	0.9515	0.8794	0.9570
9	-0.2969	-0.4613	0.9468	0.9686	0.9500
10	-0.1831	-0.2578	0.9187	0.8124	0.9186
11	-0.2176	-0.2662	0.9439	0.6693	0.9447
12	-0.1680	-0.1910	0.9008	0.7327	0.9021
13	-0.1312	-0.1401	0.6480	0.9402	0.6485
14	-0.1654	-0.2119	0.8859	0.9143	0.8861
15	-0.1623	-0.1644	0.6637	0.8037	0.6638
16	-0.2125	-0.2972	0.8915	0.8473	0.8977
17	-0.2059	-0.2659	0.8498	0.7754	0.8479
18	-0.1402	-0.1611	0.9110	0.7471	0.9133
19	-0.1490	-0.1776	0.8976	0.7811	0.9015
20	-0.1775	-0.1935	0.5911	0.9402	0.5919
21	-0.1538	-0.1578	0.6873	0.8349	0.6876
22	-0.1490	-0.1824	0.8310	0.9189	0.8298
23	-0.0963	-0.0044	0.9382	0.8205	0.9391
24	-0.2119	-0.2531	0.8254	0.8934	0.8265
25	-0.1153	-0.1149	0.8020	0.6467	0.8020
26	-0.1655	-0.1680	0.6970	0.7449	0.6980
27	-0.1234	-0.1317	0.6755	0.8371	0.6769
28	-0.0663	-0.0668	0.8951	0.4258	0.8986
avg.	-0.1563	-0.1705	0.8194	0.8054	0.8208

비교 알고리즘은 Eilersten 등의 알고리즘^[5], ExpandNet^[6], Huo 등의 알고리즘^[18], Akyüz 등의 알고리즘^[2], 총 4개의 기존 알고리즘과 비교하였다. 정량적 지표 비교 결과는 표 1과 표 2와 같다. 표 1과 표 2는 각각 HDR 정량적 화질평가 지표인 지각적 연속성 부호화 최대 신호 대 잡음비 (perceptually uniform encoding peak signal-to-noise ratio; PU-PSNR), 지각적 연속성 부호화 구조적 유사성 (perceptually uniform encoding structural similarity; PU-SSIM) 및 고동적범위 가시 차이 예측자 (high dynamic range visible difference predictor; HDR-VDP) 2.2를 사용하였다^[15,16]. 각각

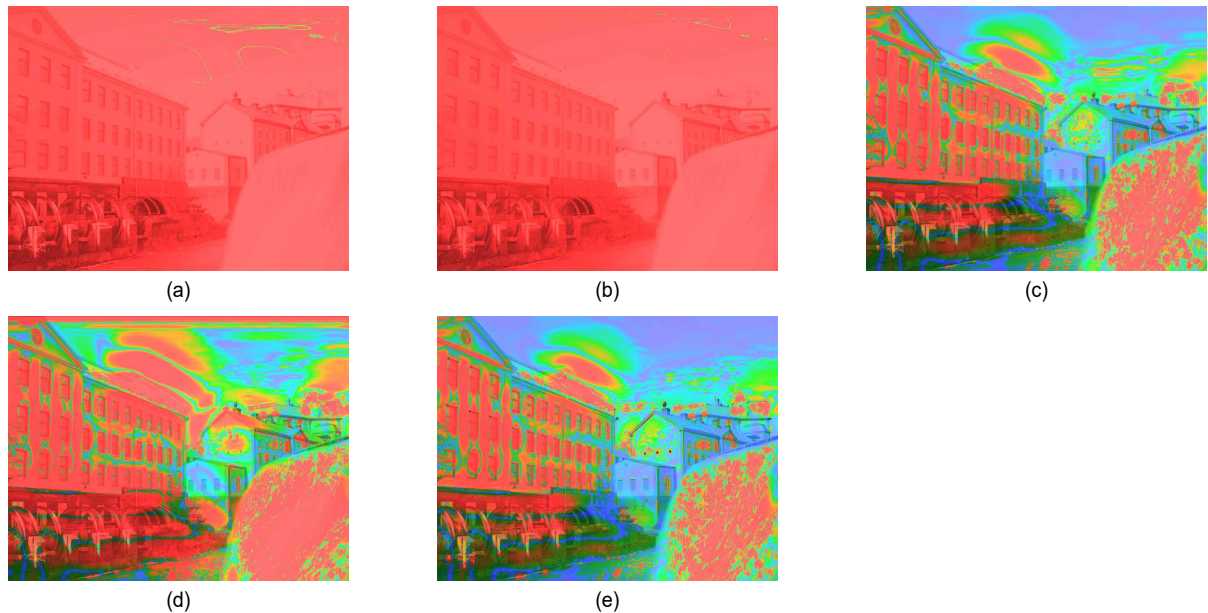


그림 3. 각 역 톤 매핑 방법들의 HDR-VDP-2.2 확률 맵 시각화 비교결과(21번 영상), (a) Akyüz 등의 알고리즘, (b) Huo 등의 알고리즘, (c) Eilersten 등의 알고리즘, (d) ExpandNet, (e) 제안하는 알고리즘

Fig 3. The comparison results of inverse tone-mapping methods(at no.21 image), (a) Akyüz et al., (b) Huo et al., (c) Eilersten et al., (d) ExpandNet, (e) proposed algorithm

PSNR, SSIM 및 VDP 지표를 HDR 영상에 적용하기 위해 제안된 정량적 화질 평가를 위한 지표이다. 각기 다른 최대 밝기값을 가지는 HDR 영상들을 균일하게 부호화하기 위해 RGB 색공간으로 변환할 때 생기는 지각적 연속성(perceptually uniform)을 이용하여 계산한다^[15].

시험 데이터셋의 영상은 모두 일부 포화되어 있기 때문에 Akyüz 등의 알고리즘과 Huo 등의 알고리즘은 모두 손실된 화소를 복원하지 못했다. 또한 지각적 연속성이 고려되지 않아 평균 PU-PSNR과 PU-SSIM이 매우 낮게 측정되었다. 반면 딥 러닝을 사용한 ExpandNet과 Eilersten 등이 제안한 알고리즘은 포화 영역의 복원값을 추정하고 해당 영역을 다시 매핑하기 때문에 원본에 가깝게 화소값을 복원하였다. PU-PSNR은 ExpandNet이 본 알고리즘보다 20번 영상에서 최대 10dB 이상 높게 나오는 경우도 있었다. 하지만 제안하는 알고리즘이 PU-PSNR 편차가 적고 28번 영상에서는 최대 9dB가 향상되었다. 결과적으로 ExpandNet보다 평균 PU-PSNR은 약 0.6dB 이상 높았으며 비교 알고리즘 중에서 가장 높게 측정되었다. 평균 PU-SSIM의 경우

ExpandNet에 비해 최소 약 0.02 이상 더 높게 측정되었는데 이는 WGIF에 의해 원본의 구조적 특징이 많이 보존되어 더 높게 측정되었다고 볼 수 있다.

그림 3은 HDR-VDP-2.2의 계산결과를 시각화하여 보여주는 결과 비교 영상이다. HDR-VDP 시각화 지표는 붉은색을 떨수록 원본과 다를 확률이 높다는 것을 의미하고 푸른색을 떨수록 원본과 같을 확률이 높다는 것을 의미한다. 21번 영상의 결과 영상을 보면 제안한 알고리즘의 결과는 ExpandNet과 거의 유사하지만 강한 하이라이트 영역에서 그 차이가 나는 것을 볼 수 있다. 또한 창문과 같이 디테일이 많이 존재하는 영역에서는 Eilersten의 알고리즘에 비해 복원 확률이 더 높게 측정되는 것을 확인할 수 있다. 이는 WGIF를 통해 에지 영역을 보존하였기 때문에 기존의 알고리즘에 비해 원본에 더 가깝게 복원되었다고 볼 수 있다.

그림 4는 각 역 톤 매핑 결과들을 다시 LDR로 톤 매핑하였을 때, 그 주관적 화질 비교 결과를 나타내었다. (g-1)은 복원된 포화 영역을 확대한 결과 영상이다. 비교 결과를 보면 기존의 사라진 하이라이트 부분이 복원된 것을 확인할

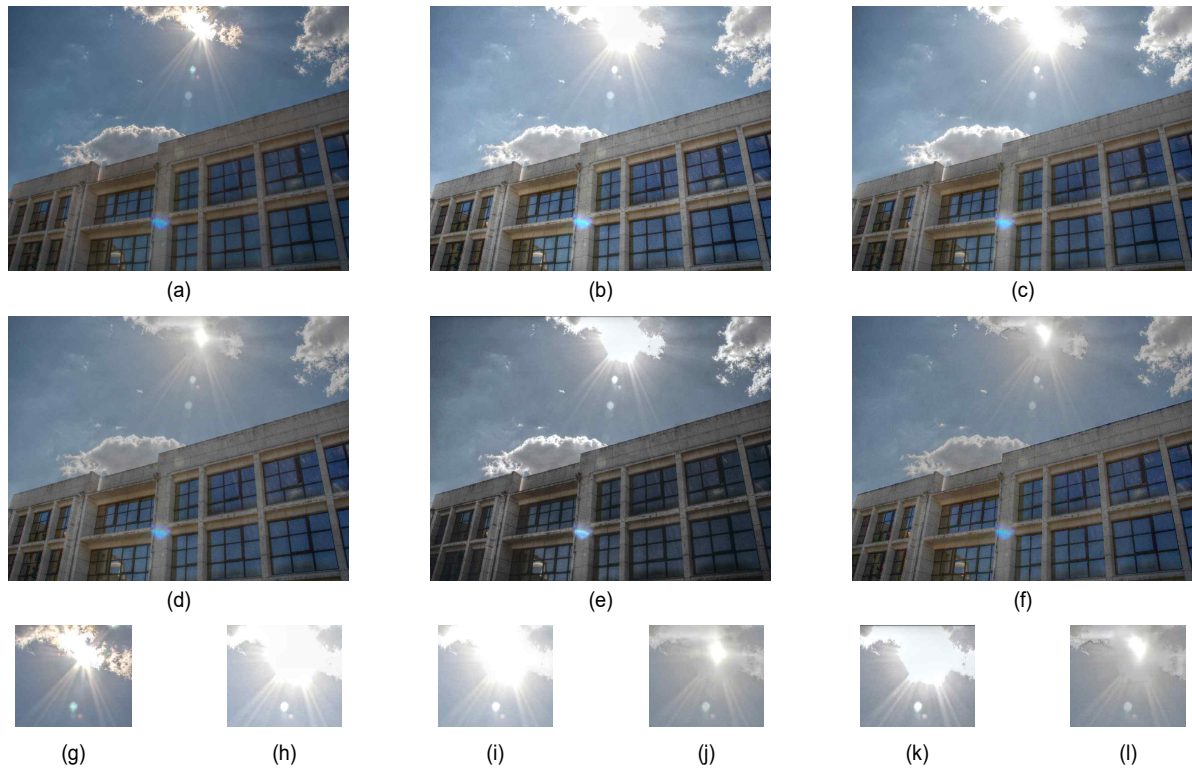


그림 4. 각 방법의 톤 매핑 주관적 화질 비교 결과(2번 영상), (a,g) 원본 영상, (b,h) Eilersten 등의 알고리즘, (c,i) ExpandNet, (d,j) Akyüz 등의 알고리즘, (e,k) Huo 등의 알고리즘, (f,l) 제안하는 알고리즘의 결과 영상과 예측된 포화 영역 영상
 Fig 4. The comparison of tone-mapping result subjective image quality of each method(at no.2 image), (a,g) ground truth, (b,h) Akyüz et al., (c,i) Huo et al., (d,j) Eilersten et al., (e,k) ExpandNet, (f,l) proposed algorithm

수 있었다. 다만 다른 알고리즘들의 경우 포화돼버린 화소 영역이 복원되지 않아 광원의 형태를 복원하지 못했다. 여기서 기존의 알고리즘에 비해 제안하는 알고리즘은 역 감마 보정을 통해 광원의 형태를 좀 더 예측하기 쉽게 보정을 해주었기 때문에 기존의 알고리즘에 비해 원본에 가깝게 광원을 복원한 것을 볼 수 있다.

V. 결론

본 논문에서는 기존의 역 톤 매핑 알고리즘들에 대한 소개와 기존의 역 톤 매핑 알고리즘의 결과를 개선하는 전처리 필터를 제시하였다. 영상의 히스토그램을 보정하는 감마 보정과 영상의 각 디테일들을 보존하고 개선시키는

WGIF를 제안하였고, 이러한 전처리 필터를 거친 영상을 역 톤 매핑 신경망에 인가하였을 때, 기존의 알고리즘보다 개선된 결과를 보여준다는 것을 HDR 정량적 화질평가를 통해 실험적으로 보였다.

많은 역 톤 매핑 알고리즘을 사용하기에 앞서 본 논문에서 제안한 해당 전처리 필터를 거친다면 좀 더 개선된 HDR 복원 영상을 얻어낼 수 있을 것이다. 또한 제안된 알고리즘은 HDR 복원하는 사업 영역에만 사용되는 것이 아니라 실험 결과에서 보인 것처럼 다시 LDR로 톤 매핑 하였을 경우 관측할 수 없었던 영역을 복원하기 때문에 사용자의 디스플레이에서 복원된 영상을 출력할 수 있다. 때문에 한정된 동적범위, 각기 다른 조명환경 등에 의해 포화되어 잃어버린 화소정보를 복원하여 영상 인식, 분석, 개선 등 다양한 산업분야에서 사용될 수 있을 것이라고 생각된다.

참 고 문 헌 (References)

- [1] H. Landis, "Production-ready global illumination," *SIGGRAPH Course Notes*, Vol.16, pp.87-101, 2002.
- [2] A. O. Akyüz, R. Fleming, B. E. Riecke, E. Reinhard, and H. H. Bühlhoff, "Do HDR displays support LDR Content?: A psychophysical evaluation," *ACM Transaction on Graphics*, Vol.26, No.38, July 2007.
- [3] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers, *Advanced High Dynamic Range Imaging: theory and practice*, A K Peters/CRC Press, New York, February 2011.
- [4] A. G. Rempel, M. Trentacoste, H. Seetzen, H. D. Young, W. Heidrich, L. Whitehead, and G. Ward, "LDR2HDR: On-the-fly reverse tone mapping of legacy video and photographs," *ACM Transaction on Graphics*, Vol.26, No.39, 2007.
- [5] G. Eilertsen, J. Kronander, G. Denes, R. Mantiuk, and J. Unger, "HDR image reconstruction from a single exposure using deep CNNs," *ACM Transactions on Graphics*, Vol.36, No.6, pp.1-15, 2017.
- [6] D. Marnerides, T. Bashford-Rogers, J. Hatchett, and K. Debattista, "ExpandNet: A Deep Convolutional Neural Network for High Dynamic Range Expansion from Low Dynamic Range Content," *Computer Graphics Forum*, Vol.37, No.2, pp.37-49, 2018
- [7] A. A. Goshtasby, "Fusion of Multi-exposure Images," *Image and Vision Computing*, Vol.23, pp. 611-618, June 2005.
- [8] S. Hecht, "The visual discrimination of intensity and the Weber-Fechner law," *The Journal of General Physiology*, Vol.7, pp.235-267, 1924.
- [9] G. E. Hinton and R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," *Science*, Vol.313, No.5786, pp.504-507, 2006.
- [10] P. Vincent, H. Larochelle, Y. Bengio and P. Manzagol, "Extracting and composing robust features with denoising autoencoders," *Proceeding of 25th International Conference on Machine Learning(ICML)*, pp.1096-1103, 2008.
- [11] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", *Arxiv.org*, 2014, <https://arxiv.org/abs/1409.1556>
- [12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceeding of IEEE conference on computer vision and pattern recognition(CVPR)*, pp.770-778, 2016.
- [13] M. D. Fairchild, "The HDR Photographic survey," *Proceeding of Color and Imaging Conference*, Vol.2007, No.1, pp.233-238, 2007.
- [14] G. Ward, "Hight Dynamic Range Image Encodings," 2006.
- [15] T. O. Aydin, R. Mantiuk, and H. P. Seidel, "Extending quality metrics to full luminance range images," *Human Vision and Electronic Imaging XIII*, Vol.6806, pp.68060B, March 2008.
- [16] M. Narwaria, R. Mantiuk, M. P. Da Silva, and P. Le Callet, "HDR-VDP-2.2: a calibrated method for objective quality prediction of high-dynamic range and standard images," *Journal of Electronic Imaging*, Vol.24, No.010501, 2015.
- [17] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," *Proceeding of the 27th International Conference on Machine Learning(ICML)*, pp.807-814, 2010.
- [18] Y. Huo, F. Yang, L. Domg, and V. Brost, "Physiological inverse tone mapping based on retina response," *The Visual Computer*, Vol.30, pp.507-517, 2014.
- [19] Z. Liand J. Zheng, Z. Zhu, W. Yao, and S. Wu, "Weighted Guided Image Filtering," *IEEE Transaction on Image Processing*, Vol.24, No.1, pp.120-129, 2015.
- [20] K. He, J. Sun, and X. Tang, "Guided Image Filtering," *Proceeding of European Conference on Computer Vision(ECCV)*, Berlin, Heidelberg, pp.1-14, 2010.
- [21] J. An, S. Lee, J. Kuk, and N. Cho, "A multi-exposure image fusion algorithm without ghost effect," *Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011.

저 자 소 개



강 라 훈

- 2018년 2월 : 한양대학교 융합전자공학부 학사
- 2018년 3월 ~ 현재 : 한양대학교 전자컴퓨터통신공학과 석사과정
- ORCID : <https://orcid.org/0000-0002-2291-2014>
- 주관심분야 : 영상처리, 딥 러닝

저 자 소 개



박 범 준

- 2016년 2월 : 한양대학교 융합전자공학부 학사
- 2016년 3월 ~ 현재 : 한양대학교 전자컴퓨터통신공학과 석박사통합과정
- ORCID : <https://orcid.org/0000-0003-3783-8272>
- 주관심분야 : 영상처리, 딥 러닝



정 제 창

- 1980년 2월 : 서울대학교 전자공학과 학사
- 1982년 2월 : KAIST 전기전자공학과 석사
- 1990년 : 미국 미시간대학 전기공학과 공학박사
- 1980년 ~ 1986년 : KBS 기술연구소 선임연구원 (디지털 및 뉴미디어 연구)
- 1990년 ~ 1991년 : 미국 미시간대학 전자컴퓨터공학부 연구교수 (영상 및 신호처리 연구)
- 1991년 ~ 1995년 : 삼성전자 HDTV 연구개발 담당 수석연구원
- 1995년 ~ 현재 : 한양대학교 전자컴퓨터통신공학과 교수 (영상통신 및 신호처리 연구실)
- ORCID : <http://orcid.org/0000-0002-3759-3116>
- 주관심분야 : 영상처리, 영상압축