

Stack-Attention을 이용한 Retinex 영상 강화 기법

박채림¹, 조석제[°], 이광일[°]¹한국해양대학교 제어계측공학과[°]한국해양대학교 제어계측공학과

cofla7572@naver.com, sjcho@kmou.ac.kr, leeki@kmou.ac.kr

Retinex image enhancement techniques
using Stack-AttentionChae-rim Park¹, Seok-je Cho[°], Kwang-il Lee[°]¹Dept. of Image Processing, Korea Maritime and Ocean University[°]Prof. of Image Processing, Korea Maritime and Ocean University[°]Prof. of Image Processing, Korea Maritime and Ocean University

요 약

광원 자체의 밝기가 낮거나 드리워진 그림자 등의 이유로 어두운 영역을 포함하고 있는 저조도 영상으로 인해 물체의 식별이 어려운 상황을 일상생활에서 겪게 된다. 본 논문에서는 조명 성분의 영향을 줄이고 객체의 특징을 표현하는 반사 성분을 강조하여 화질을 개선한다. 또한 촬영하는 카메라와 영상의 물체 사이의 상대적인 움직임으로 발생하는 흐릿한 영역을 최대한 제거해주고 잡음까지 보정이 되는 Stack-attention 기법을 제안한다.

1. 서론

최근 디지털 기술의 발달로 인해 뛰어난 화질의 영상을 얻을 수 있는 디지털 카메라들이 지속적으로 개발되고 있다. 하지만 날씨, 장소와 같은 환경적인 요인으로 인해, 디지털 카메라로 획득한 영상은 눈으로 보는 것과 다르게 대조비가 약화된다. 뿐만 아니라 카메라와 영상의 물체 사이의 상대적인 움직임으로 인한 영상 테그라데이션 현상 (degradation phenomenon)이 발생한다. 이는 카메라의 흔들림이나 물체의 역동적인 움직임으로 인해 종종 발생하게 되는데, 이러한 이유로 인해 영상 속 객체의 본연의 색을 잃게 되고 정보의 손실을 가져오는 문제가 발생하여 영상의 가독성이 떨어지게 된다.

일반적으로 Retinex[1]를 기반으로 저조도 영상 향상을 위한 여러 모델이 제안[2]되었고 SSR[3], MSR 그리고 MSR with color restoration[4] 기법이 영상의 화질 개선 기법으로 널리 사용되고 있다. 본 논문에서는 이러한 문제점을 개선하기 위해 Retinex 기반의 CNN 영상 보정 알고리즘을 제안한다. 이는 영상을 반사도와 조명 성분으로 분해한 뒤, 딥 멀티패치 계층 네트워크(deep multi-patch hierarchical network)의 수정된 인코더-디코더(encoder-decoder)를 거치게 된다. 이 과정을 통해 추출되는 잔차

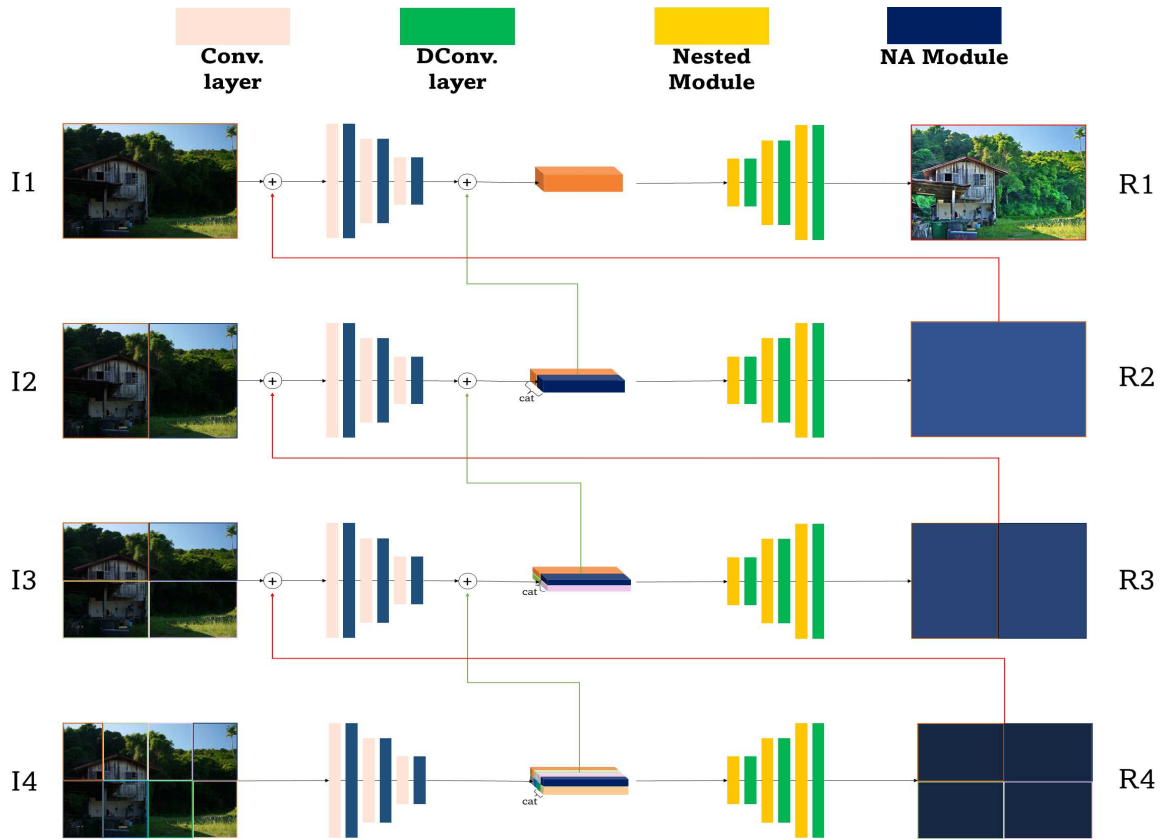
(residuals)가 출력이 된다. 이는 상위 레벨 입력으로 들어가는 과정을 반복하여 최종적으로 개선된 영상을 생성한다.

2. 제안한 방법

본 논문에서는 CNN으로 설계된 딥 멀티패치 계층 네트워크[5,6]를 기반으로 한다. 네트워크는 인코더-디코더 아키텍처를 수정하게 되는데, 이는 다음과 같다. 1) 인코더-디코더의 빌딩블록 (building-block) 모듈을 중첩 모듈로 대체하여 잔차에 가장 높은 수준의 기여를 개선할 수 있다. 2) 어텐션 메커니즘(attention-mechanism)[7,8]을 도입하여 영상 속에 있는 객체에 대한 초점을 맞출 수 있다. 이는 영상 속에 있는 모션 장면에서 다방면으로 표시된다.

2.1 중첩 어텐션 모듈

중첩 모듈은 흐릿한 영상에서의 디테일을 잡아주는 능력을 갖고 있다. 하지만 특징맵의 모든 영역이 동일하게 중요하지 않기 때문에 영상 속에서 특징마다 다른 가중치를 할당해야 한다. 그러므로 본 논문에서는 흐릿한 영상에 차별적인 특징을 추출할 수 있도록 어텐션 메커니즘을 도입하여 영상의 표현력을 높인다. 실제 어텐션 메커니즘은 채널과 공간으



(그림 1) Stack-attention 아키텍처.

로 나누어진다. 채널 어텐션은 ‘what’에, 공간 어텐션은 ‘Where’에 집중하기 때문에 이 두 개를 혼합하면 디테일한 특징을 강조할 수 있다. 이것이 중첩 모듈과 함께 순차적으로 쌓인다. 이를 통해 영상 속에서 무엇을 어디에서 강조하거나 억제하는지 학습하여 대상의 위치를 결정하고 동시에 객체 인식에 도움을 준다.

2.2 Stack-attention 네트워크 아키텍처

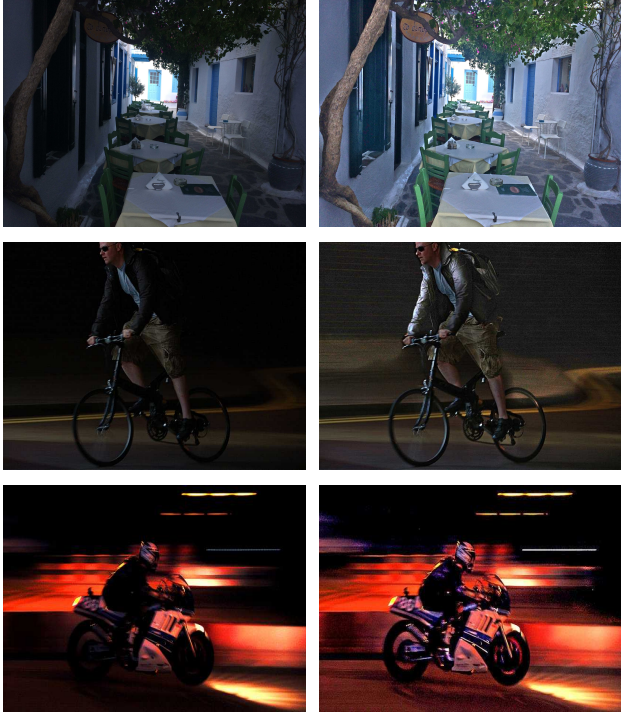
본 논문에서 제안한 전반적인 네트워크 아키텍처는 (그림 1)과 같다. 레벨 4에서 1까지 입력 영상을 각각 8, 4, 2, 1 패치로 균등하게 나눈다. 딥 멀티 계층 네트워크의 인코더 입력으로는 특정 수준의 각 패치가 독립적으로 들어가고, 디코더에 의해 잔차 기여도가 결과로 생성되고 이후 상위 레벨의 입력에 다시 추가되며 쌓이는 원리이다. 이 과정에서 각 패치의 중간 기능을 얻을 수 있다. 또한 컨볼루션 연산으로 각각의 특징 맵은 대응하는 패치와 동일한 크기를 나타낸다. 입력 시퀀스를 동일한 비중으로 참고하는 것이 아닌 예측 영상과 관련있는 픽셀을 치중해서 보는 것이다. 그리고 제안한 아키텍처에

레벨 1에서 최종적으로 추출된 영상과 Ground-truth 사이에서 발생하는 Mean Squared Error(MSE)를 최소화한다.

3. 실험

(그림 2)는 본 알고리즘을 이용한 실험 결과 영상이다. (a)는 입력 영상, (b)는 본 논문의 결과 영상이다. 입력 영상의 반사 성분과 조명 성분으로 분리한 뒤, 수정된 딥 멀티 패치 계층 네트워크를 이용하면 더욱 개선된 영상을 획득할 수 있음을 확인하였다. 실험 결과를 통해 영상 속 물체들의 색의 향상도 유지할 수 있어서 뚜렷한 객체 인식도 가능하다. 또한 역동적인 움직임으로 인해 흔들린 모션도 잡아준 것을 확인할 수 있다.





(그림 2) 실험 결과.

4. 결론

여러 환경적인 요인으로 인해 대조비가 낮아지면서 영상 내에 있는 주요 객체의 색상 및 특징 정보가 손실되기 때문에 영상 판독의 어려움이 존재한다. 본 논문에서는 어두운 영역을 포함하고 있는 저조도 영상을 제안한 방법인 Stack-attention을 이용하여 우수한 결과를 보였다.

참고문헌

- [1] E. H. Land and J. J. McCann, "Lightness and retinex theory," *Josa*, vol. 61, no. 1, pp. 1 - 11, 1971.
- [2] Fu, Q., Jung, C., and Xu, K, "Retinex-based Perceptual Contrast Enhancement in Images Using Luminance Adaptation," *IEEE Access* 6, 61277 - 61286. doi:10.1109/access.2018.2870638, 2018.
- [3] Si, L., Wang, Z., Xu, R., Tan, C., Liu, X., and Xu, J, "Image Enhancement for Surveillance Video of Coal Mining Face Based on Single-Scale Retinex Algorithm Combined with Bilateral Filtering," *Symmetry* 9 (6), 93. doi:10.3390/sym9060093, 2017.
- [4] Ma, J., Fan, X., Ni, J., Zhu, X., and Xiong, C, "Multi-scale Retinex with Color Restoration Image

Enhancement Based on Gaussian Filtering and Guided Filtering," *Int. J. Mod. Phys. B* 31 (16 - 19), 1744077. doi:10.1142/s0217979217440775, 2017.

[5] Seungjun Nah, Tae Hyun Kim, and Kyoung u Lee. "Deep multi-scale convolutional neural network for dynamic scene deblurring," In *Proc. IEEE Conf. Comp. Vis. Patt. Recogn*, pp. 257 - 265, 2017.

[6] D. Bichara, y A. Iggidr, "Multi-patch and multigroup epidemic models: a new framework," *Journal of Mathematical Biology*, vol. 77, no.1, pp. 107-134, 2017.

[7] Minh-Thang Luong, Hieu Pham and Christopher D. Manning. "Effective approaches to attention-based neural machine translation," *arXiv preprint arXiv:1508.04025*, 2015.

[8] C. Ramet, P. N. Garner, M. Baeriswyl and A. Lazaridis, "Context-Aware attention mechanism for speech emotion recognition," *IEEE Spoken Language Technology Workshop*, pp. 126-131, 2018.