

비균일 조명 환경에서의 RGB-HSI 복원을 위한 End-to-End 통합 학습 프레임워크

민웅규, 임명은, 김한빛, 정호열, 박수준
한국전자통신연구원

e-mail : {minwg, melim, coreavit, hoyoul.jung, psj}@etri.re.kr

An End-to-End Learning Framework for RGB-HSI Reconstruction in Non-Uniform Lighting Environments

Woong-Gyu Min, Myung-Eun Lim, Han-Vit Kim, Ho-Youl Jung, Soo-Jun Park
Electronics and Telecommunications Research Institute

Abstract

Reconstructing HyperSpectral Images (HSIs) from RGB inputs becomes significantly more challenging under non-uniform illumination. To address this issue, we propose an end-to-end framework that jointly performs illumination correction and spectral reconstruction within a unified network. By optimizing both stages simultaneously, the model effectively suppresses illumination-related artifacts and enhances spectral consistency. Experimental results demonstrate that the proposed framework achieves superior reconstruction performance compared to standalone RGB-to-HSI models, confirming its robustness in real-world non-uniform lighting environments.

I. 서론

초분광 영상은 가시광선뿐 아니라 적외선 등 다양한 파장 영역의 스펙트럼 정보를 포함하며, RGB 영상으로는 담아낼 수 없는 정보를 포함할 수 있기에 의료 영상 진단, 농업 감시 등 다양한 분야에서 중요한 데이터로 활용된다. 그러나 초분광 영상 획득 장비는 매우 고가이고, 데이터 처리 시간이 많이 소요되며, 현장 적용 시 장비 운영이 복잡하다는 한계가 있다.

이러한 초분광 영상 획득의 한계 극복을 위해, RGB 영상으로부터 초분광 영상을 복원하는 연구가 활발히 진행되고 있다.

RGB 영상으로부터 초분광 영상을 복원하는 문제는 메타머리즘 (Metamerism)으로 인해 비정칙 역 문제 (ill-posed Inverse Problem)로 정의된다. 서로 다른 스펙트럼 분포를 갖는 물체가 동일한 RGB 값을 생성할 수 있으므로, 입력 RGB 영상으로부터 원래의 스펙트럼을 유일하게 복원할 수 없어 해의 유일성과 안정성이 보장되지 않는 문제가 발생한다 [1].

역 문제의 비정칙 특성은 비균일 조명 환경에서 심화된다. 조명의 왜곡으로 인해 RGB 영상의 신호 대 잡음비 (Signal-to-Noise Ratio, SNR)가 저하되며, 이는 색상 정보의 왜곡으로 이어진다. 이러한 색상 정보 왜곡은 메타머리즘 공간을 확대시키고, 가능한 해의 개수를 증가시키며 복원의 안정성을 더욱 저하시킨다 [2].

본 연구는 비균일 환경에서의 강건한 RGB-초분광 복원을 위한 end-to-end 통합 학습 프레임워크를 제안한다. 제안하는 네트워크는 조명 개선 네트워크를 통해 입력 RGB 영상의 조명 환경을 균일하게 정규화하고, 초분광 영상 복원 네트워크를 통해 RGB 영상으로부터 초분광 영상을 복원한다. 전체 네트워크의 통합 학습을 통해 비균일 조명 환경에서 강건한 성능을 달성한다.

II. 초분광 영상 복원 통합 모델 설계

RGB-초분광 영상 복원 모델은 조명 개선 모델과 초분광 영상 복원 모델로 구성되며 end-to-end 학습을 통해 최적화된다. 제안하는 프레임워크의 구조는 그림 1과 같다.

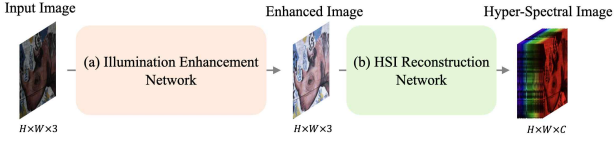


그림 1. RGB-초분광 영상 복원 통합 네트워크 구조

2.1 조명 개선 네트워크

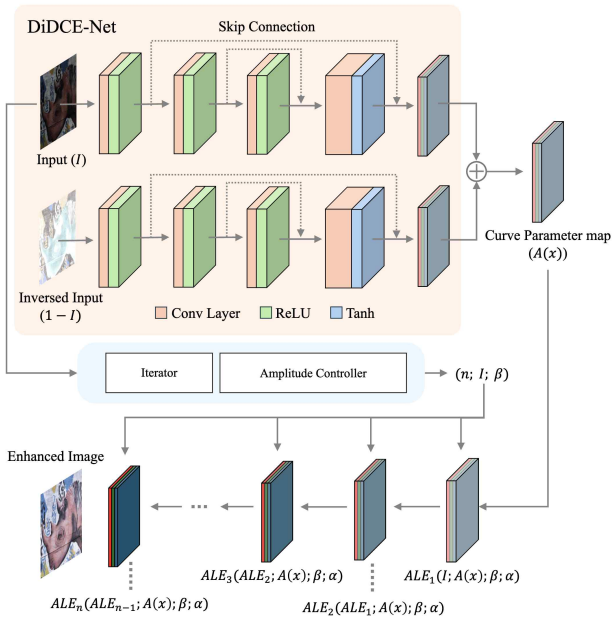


그림 2. Zero-DiDCE 기반 조명 개선 네트워크 구조

Zero-DCE(Zero-reference Deep Curve Estimation)는 비참조 기반 곡선 추정 방법으로, 참조 영상 없이도 비균일 조명 개선을 가능하게 한다. 또한 경량 네트워크 구조로 학습 부담이 적어, 초분광 영상 복원 모델과 end-to-end 학습에 유리하다. 본 연구에서는 Zero-DCE 기반 모델인 Zero-DiDCE를 활용한다 [4].

Zero-DiDCE의 네트워크 구조는 그림 2와 같다. DiDCE-Net은 원본 영상(I)과 반전된 영상($1-I$)을 동시에 입력으로 받아, 분할 비참조 손실 함수를 통해 단일 영상에서도 다양한 노출 정보를 학습한다.

2.1.1 적응형 조명 향상 곡선

DiDCE-Net은 3개의 합성곱 계층과 잔차 연결이 있는 1개의 Tanh 계층으로 구성되며, 최종적으로 곡선 매개변수 맵 $A(x)$ 를 생성한다. 이후 적응형 조명 향상 곡선 (Adaptive Light Enhancement Curve, ALE-Curve)을 반복적으로 적용하여 이미지의 조명 수준을 목표 조명값 (α)으로 수렴시킨다. ALE-Curve는 식(1)과 같다:

$$ALE_n(x) = ALE_{n-1}(x) + A(x)(ALE_{n-1}(x)^2 - ALE_{n-1}(x)) \frac{\alpha - ALE_{n-1}(x)}{\beta - ALE_{n-1}(x)} \quad (1)$$

여기서 $ALE_0(x)$ 는 입력 RGB 영상이고, $A(x)$ 는 곡선 매개변수이다. 최적 반복 횟수 n 은 입력 영상의 조명 정보에 따라 반복기(Iterator)를 통해 결정되며, 반복기는 식(2)-(3)과 같다:

$$I_{iter} = \begin{cases} -2.5x + 10, & x < 0.1 \\ 17.14x^2 - 15.15x + 10, & 0.1 \leq x < 0.45 \\ 5.66x^2 - 2.93x + 7.2, & x \geq 0.45 \end{cases} \quad (2)$$

$$n = \lfloor I_{iter} \rfloor \quad (3)$$

α 는 목표 조명값으로 이상적 노출 수준을 0.63으로 설정하며, β 는 억제 증폭 조정 곡선 의미하며 식(4)와 같이 단순하며 효과적인 곡선으로 억제 증폭을 계산한다.

$$\beta = -0.79x^2 + 0.81x + 1.41 \quad (4)$$

2.1.2 분할 비참조 손실 함수

Zero-DiDCE는 입력 영상의 국소 조명 품질에 따라 손실 정도를 구분하여 서로 다른 손실 함수를 적용한다. 단계별 비참조 손실 함수는 식(5)와 같다:

$$L = \begin{cases} W_1 L_1, & Y_i \leq Q_1 \text{ or } Y_i \geq Q_2 \\ W_2 L_2, & Y_i \in (Q_1, Q_2) \end{cases} \quad (5)$$

$$L_1 = \frac{1}{K} \sum_{e=1}^K (Y_e - E)^2 \quad (6)$$

$$L_2 = \frac{1}{K} \sum_{i=1, e=1}^K (Y_e - E)^2 \frac{1}{1 + m Y_i} \quad (7)$$

여기서 Y_i 는 국소 조명 품질, E 는 목표 조명값, Q_1, Q_2 는 극저노출과 과노출을 판단하는 임계값을 나타낸다. 영상의 국소 영역이 극도로 어둡거나 과노출된 경우, L_1 손실을 통해 강력한 조명 보정을 유도한다. 일반 조명 영역의 경우, L_2 손실을 통해 보수적인 조정을 적용한다. 이러한 분할 손실 함수를 통해 다양한 노출 수준의 영상에 효과적으로 대응할 수 있다.

2.2 초분광 영상 복원 네트워크

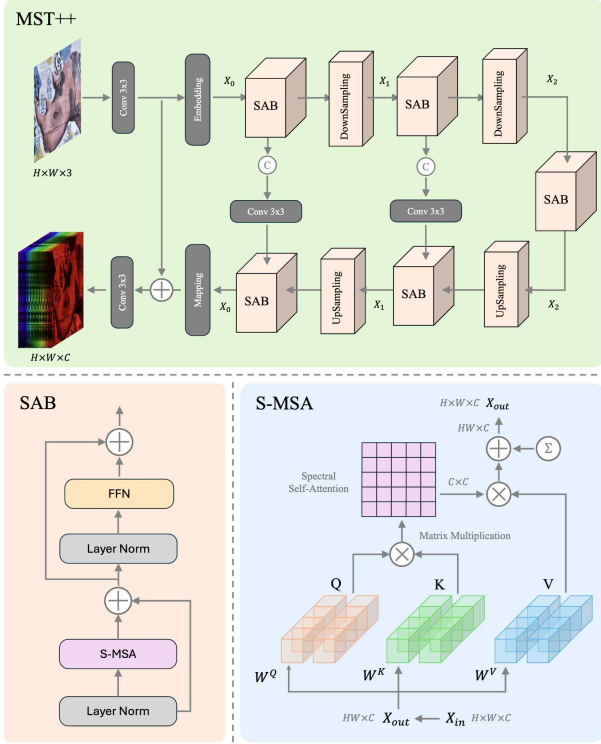


그림 3. MST++ 기반 초분광 영상 복원 네트워크 구조

MST++ (Multi-stage Spectral-wise Transformer)은 초분광 영상 복원을 위해 설계된 다중 구조 네트워크로, 점진적으로 초분광 영상 복원을 진행한다. 각 단계에서는 스펙트럼별 어텐션 메커니즘을 적용해서 스펙트럼 채널 간 상관성을 효과적으로 모델링한다 [4]. 모델의 구조는 그림 3과 같으며, 이와 같은 모델 설계로 제한된 RGB 영상 정보를 통해 효과적으로 초분광 영상을 복원할 수 있다.

MST++는 NTIRE 2022 Spectral Recovery Challenge에서 최고 성능을 기록하여 [5], 이후 초분광 영상 복원 연구에서 주요 백본 네트워크로 사용되고 있다. 본 연구 또한 MST++의 모델링 능력을 활용하며, 조명 개선 네트워크와의 통합 학습을 진행한다.

2.2.1 Spectral-wise Multi-head Self-Attention (S-MSA)

MST++은 S-MSA, Feed Forward Network (FFN)와 정규화 계층으로 이루어진 Spectral-wise Attention Block (SAB)을 네트워크 기본 단위로 사용한다. S-MSA는 스펙트럼 특징 맵을 하나의 토큰으로 간주하여 스펙트럼 차원을 따라 어텐션을 계산하며, 입력 X 는 쿼리 (Q), 키 (K), 값 (V)으로 식(8)과 같이 투영된다.

$$Q = XW^Q, K = XW^K, V = XW^V \quad (8)$$

$$A_j = \text{softmax}(\lambda_j \cdot \frac{K_j^T Q_j}{\sqrt{d_h}}), \text{head}_j = V_j A_j \quad (9)$$

여기서 $W^Q, W^K, W^V \in R^{C \times C}$ 는 학습 가능한 가중치이며 Q, K, V 를 N 개의 헤드로 분할한 후, 각 헤드 j 에서 스펙트럼 차원을 따라 식(9)를 통해 어텐션을 계산한다.

2.2.2 다단계 학습 전략

MST++는 단계별 Spectral-wise Transformer (SST) 구조를 가지며, 각 SST는 U-Net 구조로 설계되어 잔차 연결을 통해 정보 손실을 방지한다. 이후 여러 개의 SST를 순차적으로 연결하는 구조를 통해, 점진적으로 복원 품질을 향상시킨다. MST++의 학습에는 Mean Relative Absolute Error (MRAE)가 손실 함수로 사용되며 최종적으로 고품질의 초분광 영상을 복원한다.

2.3 End-to-end 통합 학습

전체 네트워크 학습을 위한 손실함수는 Zero-DiDCE의 분할 비참조 손실 함수와 MST++의 MRAE 손실 함수를 가중 조합으로 통합하였으며, 식(10)과 같다.

$$L_{total} = \lambda_0 \cdot L_{zero-didce} + \lambda_1 \cdot L_{mst} \quad (10)$$

여기서 $L_{zero-didce}$ 와 L_{mst} 는 각각 Zero-DiDCE의 분할 비참조 손실 함수와 MST++의 MRAE 손실 함수이며, λ_0 와 λ_1 은 손실 함수 가중 조합을 위한 가중치를 의미한다. 제안된 구조는 조명 보정 결과가 이후 단계의 스펙트럼 복원 과정에 직접 반영되며, 복원 손실이 역전파되어 조명 보정 모듈의 파라미터까지 함께 최적화되는 학습 특성을 가진다.

III. 실험 및 결과

3.1 실험 환경 및 학습 설정

3.1.1 하드웨어 환경 및 학습 파라미터

실험은 NVIDIA H100 GPU 환경에서 Pytorch 2.5.1 프레임워크를 활용하였다. 모델 학습은 300개의 epoch로 진행하였고, 배치 크기는 8로 설정하였다. 최적화 알고리즘은 Adam, 학습률은 0.0001로 초기화하였다.

평가 지표로는 초분광 영상 복원에서 주로 활용되는 평균 상대 절대 오차 (Mean Relative Absolute Error, MRAE), 평균 제곱근 오차 (Root Mean Squared Error, RMSE), 최대 신호 대 잡음비 (Peak Signal-to-Noise Ratio, PSNR)를 활용한다.

3.1.2 데이터 전처리 및 증강

실험에는 NTIRE 2022 Spectral Recovery Challenge의 ARAD-1K 데이터셋을 사용한다. ARAD-1K 데이터셋은 400-700nm 파장대의 31개 스펙트럼 채널을 포함하며, 512x512 영상 해상도를 가진다.

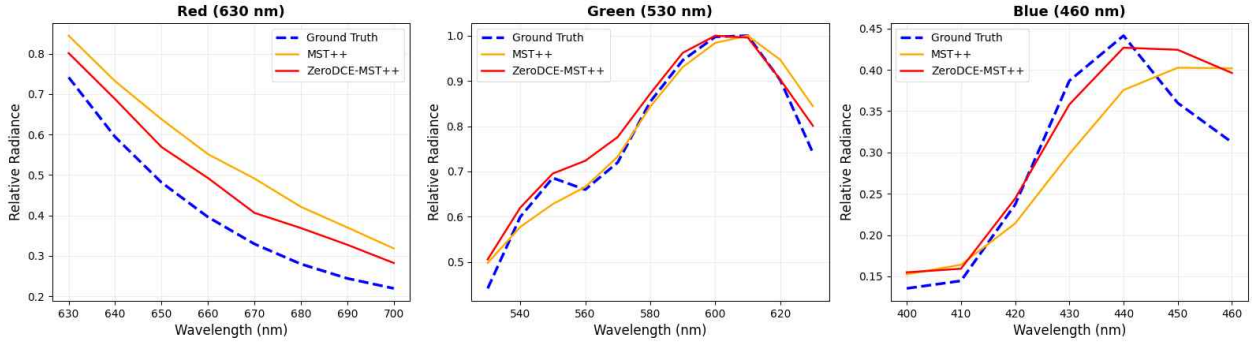


그림 4. 초분광 영상 복원 대역별 상대 방사휘도 비교 결과

데이터는 총 1000개이며 이 중 학습 검증 테스트 데이터 분할 비율은 16:1:1로 설정하였다. 입력 RGB 영상은 0-1 범위로 정규화하며, 초분광 영상은 L2 정규화를 적용한다. 상하 반전, 좌우 반전, 회전 증강을 통해 일반화 성능을 향상하였으며, 128x128크기의 패치로 나누어 패치 단위별로 학습을 진행한다.

3.2 실험 결과

표1은 제안한 복원 방법과 기존 방법의 복원 성능을 평가한 표이다. 표1의 결과, 조명 개선 네트워크를 통합을 통해 MRAE 0.1703, RMSE 0.0255, PSNR 34.396을 기록하여 비교 방법 대비 MRAE 3.34%, RMSE 2.7%의 개선을 달성하였다. 이는 조명 보정 네트워크가 입력 RGB 영상의 조도 비균일성을 제거하여 복원 모듈이 보다 안정적으로 스펙트럼 패턴을 학습한 결과로 볼 수 있다.

그림 4는 Red, Green, Blue 3가지 스펙트럼 채널에 대한 복원 상대 방사 휘도를 비교한 그래프이다. 모든 스펙트럼 채널에서 기존 MST++ 대비 복원 성능이 개선되었고, 특히 장파장 대역 (630nm)에서 성능 개선이 두드러진 것을 확인하였다. 이는 비균일 조명 환경에서 장파장의 신호 손실이 상대적으로 크고, 조명 개선이 이러한 손실을 효과적으로 보정한다는 것을 의미한다.

표 1. 초분광 영상 복원 실험 결과

	MRAE	RMSE	PSNR
MST++	0.1703	0.0255	34.396
Zero-DiDCE-MST++	0.1646	0.0248	34.269

IV. 결론 및 향후 연구

본 연구는 RGB-초분광 영상 복원을 위해 RGB 영상 조명 개선과 초분광 영상 복원 네트워크를 end-to-end 방식으로 통합 학습하는 프레임워크를 제안하였다.

제안한 방법은 기존 MST++ 모델 대비 MRAE를 3.34% 감소시키며, 400-700 nm의 모든 스펙트럼 대역에서 유의한 성능 향상을 달성하였다. 이는 조명 개선을 전처리 네트워크로 활용할 경우, 입력 RGB 영상의 신호-잡음비를 개선하여 비균일 조명 환경에서의 초분광 영상 복원 성능을 효과적으로 향상시킬 수 있음을 나타낸다. 향후에는 노이즈 제거 등 추가 전처리 작업을 통해 더욱 강건하고 범용적인 모델을 개발하고자 한다.

Acknowledgement

이 논문은 한국전자통신연구원 연구운영비지원사업의 일환으로 진행하였음 (24ZR1200, DNA 기반 국가 지능화 핵심기술개발)

참고문헌

- [1] Zhang, Jingang, et al. "A survey on computational spectral reconstruction methods from RGB to hype rspectral imaging." Scientific reports 12.1, 2022.
- [2] Hu, Yanan, et al. "Lighten-MST: Low-light Spectral Reconstruction via Illumination-guided Spectral-aware Transformer." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2025.
- [3] Mi, Aizhong, et al. "Rethinking zero-DCE for low-light image enhancement." Neural Processing Letters 56.2 (2024): 93.
- [4] Cai, Yuanhao, et al. "Mst++: Multi-stage spectral-wise transformer for efficient spectral reconstruction." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, p.745-755, 2022.
- [5] Arad, Boaz, et al. "Ntire 2022 spectral recovery challenge and data set." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, p. 863-881, 2022.