

버스트 이미지 복원을 위한 비균일 노출 예측 파이프라인[†]

이영기[°] 김우혁 조성현^{*}

포항공과대학교

{yongkee110, woohyeok, s.cho}@postech.ac.kr

Non-uniform Exposure Prediction Pipeline for Burst Image Restoration

Yongkee Lee[°] Woohyeok Kim Sunghyun Cho^{*}

Pohang University of Science and Technology (POSTECH), South Korea

요약

버스트 이미지 복원(burst image restoration)은 짧은 시간 동안 여러 장의 이미지를 연속으로 촬영하고 합성하여 한 장의 고품질 이미지를 복원하는 기술이다. 다양한 연구가 제안되었지만 대부분 균일(uniform) 노출 설정을 사용하거나, 비균일(non-uniform) 노출 설정을 사용하더라도 고정된 배율의 노출 브래킷(exposure bracket) 등 사전에 정의된 노출 설정에 의존한다. 이럴 경우, 버스트 이미지들이 서로 유사한 정도의 블러(blur)와 노이즈(noise)를 가지고 있어 상호보완적인 정보가 부족하거나 촬영 환경을 충분히 반영하지 못한다는 한계가 있다. 이러한 문제를 해결하기 위해 본 연구에서는 복원 과정에서 촬영 환경을 반영할 수 있는 비균일 노출 예측 파이프라인을 제안한다. 본 연구의 파이프라인은 주어진 환경을 고려하여 버스트 이미지 복원에 특화된 촬영이 가능하므로 저조도 환경에서의 촬영 품질을 획기적으로 향상시킬 수 있다.

Abstract

Burst image restoration is a technique that synthesizes multiple images captured consecutively over a short time period to restore a single high-quality image. While numerous studies have been conducted in this field, most rely on uniform exposure settings or predefined non-uniform configurations (e.g., fixed-ratio exposure brackets). In such scenarios, burst images often exhibit similar levels of blur and noise, leading to insufficient complementary information and a failure to adequately capture the characteristics of the imaging environment. To address these limitations, we propose a non-uniform exposure prediction pipeline that dynamically optimizes exposure parameters based on scene conditions and restoration requirements. The proposed pipeline facilitates environment-aware burst image capture, specifically tailored for restoration tasks, thereby significantly enhancing image quality in low-light conditions.

키워드: 버스트 이미징, 저조도 이미지 복원

Keywords: Burst Imaging, Low-light Image Restoration

1 서론

버스트 이미징(burst imaging)은 디지털 카메라에서 짧은 시간 동안 여러 장의 이미지를 연속으로 촬영하는 기술이다. 버스트 이미지 복원(burst image restoration)은 저조도(low-light) 환경에서 발생하는 노이즈(noise)가 각 이미지마다 무작위로 다르게 나타

난다는 점에 착안해, 버스트 이미징으로 촬영한 여러 장의 이미지를 합성하여 한 장의 고품질 이미지를 복원하는 기법이다. 버스트 이미지 복원은 오래전부터 꾸준히 연구되어 왔으며, 오늘날 많은 카메라에서 핵심적인 기술로 자리 잡았다.

버스트 이미지 복원 연구 [1, 2, 3, 4, 5, 6, 7, 8, 9, 10] 가 활발히 진행되어 저조도 환경에서의 디노이징(denoising)과 디블러링(deblurring) 성능이 획기적으로 개선되었으나 기존의 연구는 여

[†] 학부생 주저자 논문임.

*corresponding author: Sunghyun Cho / Pohang University of Science and Technology (POSTECH) (s.cho@postech.ac.kr)

전히 한계점을 가지고 있다. 기존 방법들은 이미 촬영된 버스트 이미지들을 어떻게 복원할 것인가에 초점을 맞추며, 버스트 이미지들을 어떻게 촬영할 것인지에 대해서는 충분히 고려하지 않았다. 예를 들어, 기존 방법들은 대부분 균일 노출(uniform exposure) 설정을 가정하는데, 이럴 경우 모든 버스트 이미지들이 동일한 노출 시간(exposure time)과 게인(gain)으로 촬영되고, 서로 유사한 정도의 노이즈와 블러를 가지게 되어 상호보완적인 정보의 부족으로 성능이 제한될 수 있다.

이러한 문제를 해결하기 위해 최근 비균일 노출(non-uniform exposure) 설정을 사용하는 방식 [11]이 제안되었지만 고정된 배율의 노출 브래킷(exposure bracket)과 같이 여전히 사전에 정의된(pre-defined) 노출 설정에 의존한다는 문제가 있다. 노출 설정이 사전에 정의될 경우, 주어진 환경을 충분히 반영하지 못한 채 버스트 이미지들이 촬영되기 때문에 최적의 결과를 보장하지 못한다는 근본적인 한계가 존재한다. 예를 들어, 극심한 저조도 환경에서 촬영을 한다고 가정해보자. 카메라나 물체의 움직임이 없음에도 불구하고 사전에 정의된 노출 시간이 짧은 경우, 촬영된 버스트 이미지에 포함된 많은 노이즈로 인해 복원 성능이 급격히 저하될 수 있다. 반대로 카메라나 물체의 움직임이 매우 큰 상황에서 사전에 정의된 노출 시간이 충분히 짧지 않은 경우, 흐릿하게 촬영된 버스트 이미지로 인해 마찬가지로 복원 품질이 떨어질 수 있다.

이에 본 연구에서는 주어진 환경과 복원 과정을 고려하여 촬영을 수행하는 것의 중요성을 검증하며 버스트 이미지 복원을 위한 비균일 노출 예측 파이프라인을 제안한다. 본 연구의 파이프라인은 노출 예측 모듈, 버스트 합성 모듈, 복원 네트워크로 구성되며, 실제로 적용될 때에는 버스트 합성 모듈이 실제 촬영 시스템으로 대체된다. 제안하는 파이프라인을 사용할 경우, 촬영 환경을 고려하여 버스트 이미지 복원에 특화된 버스트 이미지 촬영이 가능하기 때문에 저조도 환경에서의 촬영 품질을 획기적으로 개선할 수 있음을 확인하였다. 추가적으로 본 연구에서는 각 노출 설정의 효과에 대해 자세히 분석하고 효율적인 학습 방식도 함께 제안한다.

2 관련 연구

2.1 버스트 이미징

버스트 이미징 기법은 짧은 시간 동안 연속적으로 여러 장의 이미지를 촬영한 뒤, 이들을 효과적으로 결합하여 이미지의 품질을 크게 향상시킬 수 있는 촬영 기법이며 이미지 복원(restoration)과 개선(enhancement) 분야에 폭넓게 활용되고 있다 [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]. 이 기술은 각 버스트 프레임마다 발생하는 노이즈가 무작위로 다르게 나타난다는 점을 이용해 여러 이미지를 통합함으로써 단일 이미지 촬영에서는 억제하기 어려운 노이즈를 효과적으로 감소시킬 수 있다 [1, 2]. 또한, 피사체의 움직임이나 손떨림 등으로 인해 발생하는 블러(blur) 현상 역시 다중 프레

임 간의 정보 보안을 통해 줄일 수 있으며 결과적으로 더 선명한 이미지를 획득할 수 있다 [3, 4, 5, 6]. 특히, 버스트 이미징은 초해상화(super-resolution) 분야에서도 중요한 역할을 하는데, 여러 장의 저해상도 이미지를 정밀하게 정렬하고 융합함으로써 단일 프레임만으로는 얻을 수 없는 고해상도 이미지를 획득할 수 있다 [7, 8, 9, 10].

이렇게 버스트 이미징 기법은 저조도 환경, 빠른 움직임이 있는 장면, 혹은 촬영 조건이 불안정한 상황 등 다양한 환경에 효과적으로 적용될 수 있으며 딥러닝 기반의 정렬 및 융합 알고리즘이 발전함에 따라 그 활용 범위가 더욱 확대되고 있다. 하지만 기존의 버스트 이미징 기법들은 대부분 균일한 노출(uniform exposure) 설정을 가정하기 때문에 버스트 이미지들 간의 상호보완적인 정보를 보다 효과적으로 활용할 수 없다는 한계가 존재한다.

2.2 비균일 노출 설정을 활용한 이미지 복원

비균일 노출로 다중 이미지를 촬영하는 기법은 HDR(high dynamic range) 이미징 분야에서 오랜 기간 널리 활용되어 왔으며 다양한 노출 조건에서 촬영된 여러 이미지를 결합함으로써 장면의 넓은 명암 범위를 효과적으로 표현할 수 있다 [12, 13, 14, 15, 16, 17, 18, 19, 20, 21]. HDR 이미징뿐만 아니라 비균일 노출 설정을 다양한 영상 복원 및 개선 작업에 적용하려는 연구가 활발하게 이루어지고 있다.

이중 노출 기법(dual exposure methods)은 장노출과 단노출 이미지를 동시에 혹은 연속적으로 획득한 다음, 장노출 이미지는 적은 노이즈와 풍부한 색상 정보를 제공하고 단노출 이미지는 움직임에 의한 블러 발생을 억제하여 선명한 경계와 세부 정보를 확보할 수 있도록 한다 [22, 23, 24, 25, 26, 27, 28]. 이러한 방식은 상호보완적인 정보를 결합함으로써 노이즈 제거와 블러 제거 성능을 동시에 향상시킬 수 있다. 최근 다양한 복원 작업을 동시에 수행하기 위해 노출 브래킷(exposure bracket)을 활용하는 방법 [11]이 제안되었으며 디블러링(deblurring), HDR, 그리고 초해상화 등에서 큰 성능 향상을 보였다. 하지만 여전히 고정된 노출 설정을 사용하여 주어진 환경을 충분히 반영하지 못한 채 촬영이 수행되기 때문에 복원 성능에 한계가 존재한다.

3 방법

3.1 이미지 열화 모델

이미지 I 의 노출 e 는 노출 시간 t 와 게인 g 에 의해 결정된다. 구체적으로 e, t, g 는 다음의 관계식을 따른다:

$$e = \frac{t \cdot g}{k} \quad (1)$$

여기서 k 는 특정 카메라에 종속된 상수이고 노출 e 는 이미지 I 의 전체 밝기를 결정하며 노출 시간 t 와 게인 g 는 각각 블러와 노이

스에 영향을 미친다. 구체적으로, 카메라 RAW 공간에서 t 와 g 가 이미지 I 에 미치는 영향은 다음과 같은 이미지 열화 모델(image degradation model)로 표현할 수 있다:

$$I = \text{quant} \circ \text{cfa} \left(g \left(\int_{\tau=0}^t S_{\tau} d\tau + N_{\text{shot}} + N_{\text{read}} \right) \right) \quad (2)$$

여기서 S_{τ} 는 시간 τ 에서의 카메라 RAW 색공간 내 장면 복사조도(scene radiance)이며 N_{shot} 과 N_{read} 는 각각 샷 노이즈(shot noise)와 리드 노이즈(read noise)를 나타낸다. $\text{quant}(\cdot)$ 과 $\text{cfa}(\cdot)$ 는 각각 양자화(quantization)와 컬러 필터 배열 샘플링(color filter array, CFA)을 의미하며 \circ 는 함수 합성(function composition) 표현이다.

앞서 언급한 샷 노이즈는 빛이 광자로 구성되어 있으며 이 광자가 센서에 확률적으로 도달함에 따라 자연스럽게 발생하는 통계적 노이즈이다. 광자의 도달은 이산적이고 독립적인 사건으로 간주되며 동일한 조건에서도 각 픽셀에 도달하는 광자 수에는 차이가 발생한다. 이러한 특성은 포아송 분포(Poisson distribution)로 모델링할 수 있으며 특히 저조도 환경에서는 광자 수 자체가 적어 불규칙성이 크게 나타난다. 리드 노이즈는 센서에서 광자 신호를 디지털 신호로 변환하는 과정에서 발생하는 전자적 노이즈이다. 이는 정규 분포(Gaussian distribution)로 모델링되며 신호의 세기와 무관하게 일정한 분산을 가지므로 신호가 약한 저조도 조건에서 상대적으로 더 크게 나타난다.

3.2 비균일 노출 예측 파이프라인



Figure 1. Overview of our pipeline. During inference, the burst synthesis module is removed, and the restoration network processes real burst images captured by real camera system.

3.2.1 노출 예측 모듈

본 연구의 핵심은 촬영 환경과 복원 과정을 고려하여 적절한 버스트 이미징을 수행하는 것이다. 이를 위해 본 연구에서는 두 장의 프리뷰 이미지($I_t, I_{t-1} \in \mathbb{R}^{3 \times H \times W}$)와 I_t 를 촬영할 때의 게인 g_p 를 입력으로 받아 버스트 이미지의 노출 시간을 예측하는 ‘노출 예측 모듈’을 제안하며, 해당 모듈은 기존의 옵티컬 플로우 추정 네트워크 RAFT [29]와 단순한 형태의 2차식으로 구성된다. 학습시에는 예측된 노출 시간이 버스트 합성 모듈의 입력으로 사용되어 사실적인 버스트 이미지를 합성하고, 실제 적용시에는 예측된 노출 시간으로 실제 버스트 이미징을 수행한다.

구체적으로, 먼저 두 장의 프리뷰 이미지로부터 RAFT 네트워크를 사용해 각 픽셀의 옵티컬 플로우 벡터(optical flow vector)를 구하고 벡터의 크기를 평균 낸 다음 0과 1사이의 값으로 정규화

하여 모션 정보 m_p 를 다음과 같이 구한다:

$$m_p = \frac{1}{\epsilon} \cdot \min \left(\frac{1}{H \times W} \sum_{i=1}^{H \times W} \|\mathbf{f}_i\|_2, \epsilon \right) \quad (3)$$

여기서 $\mathbf{f}_i = (dx_i, dy_i)$ 는 RAFT를 통해 추정된 i 번째 픽셀의 옵티컬 플로우 벡터이며 H 는 이미지의 높이(수직 방향 픽셀 수), W 는 이미지의 너비(가로 방향 픽셀 수)이다. 또한, 실험에서 ϵ 는 10으로 설정하였다. 이어서 모션 정보 m_p 와 게인 g_p 를 선형 회귀를 통해 정의된 이차식에 대입하여 첫번째 버스트 이미지의 노출 시간 t_1 과 배율 r 을 구하고 최종적으로 n 개의 버스트 이미지 노출 시간 $\{t_1, t_1 \cdot r, t_1 \cdot r^2, t_1 \cdot r^3, \dots\}$ 을 획득한다. 이때, 입력으로 사용되는 게인 g_p 도 최댓값을 활용해 0과 1사이로 정규화된 값이며 선형회귀를 통해 이차식을 최적화 하는 과정은 4.1에서 자세히 설명한다.

3.2.2 버스트 합성 모듈



Figure 2. Detailed architecture of the burst synthesis module.

버스트 합성 모듈은 연속적인 선형(linear) sRGB 이미지들과 노출 예측 모듈을 통해 구한 n 개의 노출 시간을 입력으로 받아 사실적인 n 개의 버스트 이미지를 합성한다(Figure 2). 이 모듈은 블러 합성(blur synthesis), 역 표준 연산(inverse canonical operations), 그리고 노이즈 합성(noise synthesis) 단계로 구성되며 순차적으로 수행된다.

블러 합성 본 연구에서는 보다 사실적인 블러를 합성하기 위해 기존 연구 [30, 31]에서 흔히 사용되는 프레임 평균화(frame averaging) 기법을 사용한다. 구체적으로, 노출 시간 만큼의 연속적인 선형 sRGB 이미지들을 더한 다음 노출 시간으로 나누는 방식으로 블러를 합성하였으며 n 개의 노출 시간에 순서대로 적용해 n 개의 버스트 이미지를 합성하였다. 이때, 실제 촬영 시스템을 사실적으로 모사하기 위해 버스트 이미지들 간에 어느 정도의 시차가 생기도록 하였다.

역 표준 연산 역 표준 연산(inverse canonical operations)은 ParamISP [32]의 CanoNet 구조를 차용하여 역 색공간 변환(inverse color space transform), 역 화이트 밸런스(inverse white balance), 모자이크(mosaic) 순서로 적용된다. 블러 합성을 통해 얻어진 블러 이미지를 $I_{\text{blur}} \in \mathbb{R}^{3 \times H \times W}$ 라 하면 I_{blur} 에 역 색공간 변환 행렬 $M_{\text{Cam}}^{-1} \in \mathbb{R}^{3 \times 3}$ 을 적용하여 I_{blur} 의 색공간을 선형

sRGB 공간에서 RAW 색공간으로 변환한다. 이어서 각 RGB 채널을 해당 계수 $g_{WB} \in \mathbb{R}^3$ 로 나누는 역 화이트 밸런스를 수행한 후, 최종 이미지 $I_{RAW} \in \mathbb{R}^{1 \times H \times W}$ 를 얻기 위해 모자이크 연산을 적용한다.

노이즈 합성 노이즈 합성은 RAW 이미지에서 샷 노이즈(shot noise)와 리드 노이즈(read noise)가 주요한 노이즈 성분이라는 가정 하에 기존 연구 [33, 31, 11]와 유사한 방식을 따른다. 이 두 성분의 조합은 이분산 가우시안 확률 변수(heteroscedastic gaussian random variable) \mathbf{N} 으로 모델링되며 다음과 같은 수식으로 표현된다:

$$\mathbf{N} \sim \mathcal{N}(0, \lambda_{\text{read}} + \lambda_{\text{shot}} \mathbf{I}_{\text{RAW}}) \quad (4)$$

여기서 \mathbf{I}_{RAW} 는 역 표준 연산을 통해 구해진 RAW 색공간 이미지이며 λ_{read} 와 λ_{shot} 은 노이즈 파라미터이다. 노이즈 파라미터는 기존 연구 [33]와 유사하게 ISO 값과 노이즈 파라미터 간의 선형 관계를 가정하여 추정하였다.

최종적으로 샘플링된 노이즈 \mathbf{N} 을 계인 g_i 를 고려하여 \mathbf{I}_{RAW} 에 더한 다음 픽셀 값이 0과 1 사이를 벗어나지 않도록 클리핑(값 제한)하여 버스트 이미지를 합성한다. 이때, 장면 복사조도(scene radiance)는 증폭하지 않고 계인 g_i 에 따라 노이즈 \mathbf{N} 만 증폭하며 이는 본 프레임워크가 입력 버스트 이미지들의 계인이 각 이미지의 노출 시간에 반비례하도록 가정하기 때문이다. 이로 인해 모든 버스트 이미지의 밝기 수준은 동일하게 유지되지만 노이즈 수준은 달라진다.

3.2.3 복원 네트워크

본 연구에서는 버스트 이미지 복원 및 개선 분야에서 최신 모델인 Burstormer [2]를 복원 네트워크로 사용했다. 이 모델은 트랜스포머 기반 구조를 가지고 있기 때문에 여러 장의 버스트 이미지를 입력받아 시간적·공간적으로 중요한 특징을 추출할 수 있어 좋은 복원 성능을 보인다. 학습시에 복원 네트워크는 버스트 합성 모듈에서 합성된 n 개의 버스트 이미지를 입력으로 받아 노이즈와 블러가 없는 깔끔한 이미지 한 장을 복원한다.

4 실험

4.1 구현 세부사항

4.1.1 데이터셋

본 연구의 파이프라인을 학습시키기 위해 장면 복사조도 시퀀스(scene radiance sequences)로 구성된 학습 데이터셋 \mathcal{D} 를 사용한다. 즉, $\mathcal{D} = \mathbf{S}_1, \mathbf{S}_2, \dots$ 이며 여기서 \mathbf{S}_i 는 하나의 장면 복사조도 시퀀스이다. 본 연구에서는 GoPro 데이터셋 [30]의 영상 클립을 구성하는 이미지들을 장면 복사조도 시퀀스로 사용하였다.

영상 클립은 조리개값이 고정된 GoPro 카메라를 사용해 240 FPS의 고주사율(high frames per second)로 촬영되어 모션 블러가 거의 없으며, 밝은 환경에서 낮은 계인 값으로 촬영되어 노이즈 또한 매우 적다. 따라서 노이즈와 블러가 거의 없는 고품질의 정답(ground-truth) 이미지를 확보할 수 있다는 장점이 있어 해당 데이터셋을 채택하였다. 구체적으로, 각 영상 클립의 이미지들에 대해 감마 확장(gamma expansion)을 적용하여 색공간을 선형 sRGB 공간으로 변환하였다. 그런 다음, 기존의 프레임 보간(frame interpolation) 기법 [29]을 적용해 프레임 속도를 8배 증가 시킴으로써 최종적으로 1920 FPS의 영상 클립을 확보하였다.

데이터셋 \mathcal{D} 를 $\mathcal{D}_{\text{restore}}$, $\mathcal{D}_{\text{exposure}}$, 그리고 $\mathcal{D}_{\text{test}}$ 로 분할하였으며 각각 4800, 1036, 532 개의 시퀀스로 구성된다. GoPro 데이터셋의 학습 및 평가 데이터셋은 각각 22개, 11개의 영상 클립으로 합성되는데, 22개의 학습 영상 클립을 11개씩 임의로 나누어 $\mathcal{D}_{\text{restore}}$ 와 $\mathcal{D}_{\text{exposure}}$ 를 구성하고 11개의 평가 영상 클립으로 $\mathcal{D}_{\text{test}}$ 를 구성하였다. 또한, 각 시퀀스에 대해 프리뷰 이미지의 계인 g_p , 역 색공간 변환 행렬 M_{Cam}^{-1} , 화이트 밸런스 계인 g_{WB} 는 무작위로 샘플링하였으며 프리뷰 이미지의 노출 시간 t_p 은 1/120초로 설정하였다.

4.1.2 학습 및 평가 방법

본 연구의 모든 실험에서 버스트 이미지의 개수 n 은 4로 설정하였으며, 다양한 노출 시간에 대해 학습한 복원 네트워크를 동일하게 사용하여 데이터셋 $\mathcal{D}_{\text{test}}$ 의 시퀀스로 평가하였다. 학습에 대한 자세한 사항은 다음과 같다.

복원 네트워크 노출 예측 모듈을 최적화하기 위해서는 다양한 노출 시간에 강인한 복원 네트워크가 필요하기 때문에 먼저 $\mathcal{D}_{\text{restore}}$ 를 사용하여 복원 네트워크를 학습하였다. 이때, i 번째 버스트 이미지의 노출 시간 t_i 를 최소·최대 노출 시간 $[t_{\min}, t_{\max}]$ 범위에서 무작위로 샘플링하였다. 이후 모든 버스트 이미지가 동일한 밝기를 갖도록 $g_i = t_p \cdot g_p / t_i$ 공식으로 계인 g_i 를 설정하였으며 해당 장면 복사조도 시퀀스 \mathbf{S} 와 노출 시간 t_i , 계인 g_i 를 버스트 합성 모듈에 입력하여 버스트 이미지를 합성하였다. 최종적으로 합성된 n 개의 버스트 이미지를 복원 네트워크의 입력으로 사용하여 복원된 이미지를 출력하고 정답 이미지와의 평균 절대 오차(mean absolute error)를 손실 함수로 사용하여 복원 네트워크를 학습하였다.

복원 네트워크는 PyTorch [34] 프레임워크에서 사용하였으며 에포크(epoch)는 500, 학습률(learning rate)은 3.0×10^{-4} 에서 점진적으로 감소하여 1.0×10^{-8} 이 되도록 설정하였다. 또한, AdamW [35] 옵티마이저와 코사인 어닐링 스케줄러 [36]를 사용하였다. 학습에 사용된 모든 이미지는 256×256 해상도를 가지며, 4개의 GeForce RTX 3090 GPU가 장착된 PC에서 배치 크기 4로

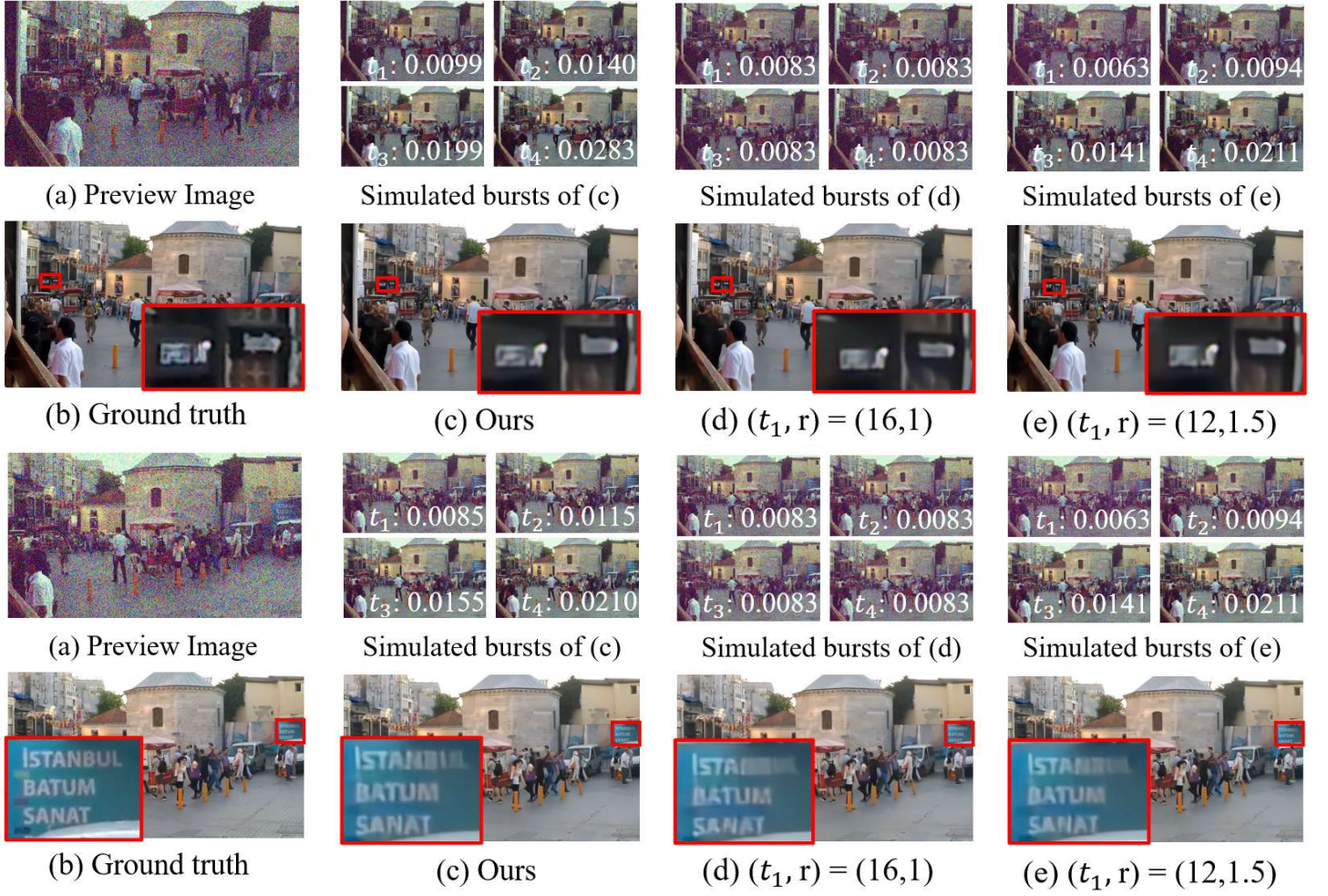


Figure 3. Qualitative results on our test set. The exposure times t_i used in (c) are predicted by our exposure prediction module, which takes the preview image (a), motion m_p and gain g_p as input information. The exposure times t_i used in (d) and (e) are calculated as $\{t_1, t_1 \cdot r, t_1 \cdot r^2, t_1 \cdot r^3\}$, using their combination (t_1, r) . (d)’s combination is $(t_1, r) = (16, 1)$ and (e)’s combination is $(t_1, r) = (12, 1.5)$.

학습을 수행하였다.

노출 예측 모듈 촬영 환경을 고려하여 복원에 적합한 버스트 이미지의 노출 시간을 예측하는 노출 예측 모듈을 학습하기 위해 먼저 앞서 학습한 복원 네트워크를 활용하여 $\mathbf{t}_{\text{pseudo-gt}}$ 를 설정한다. $\mathbf{t}_{\text{pseudo-gt}}$ 는 미리 정의된 다양한 노출 시간 조합 집합을 모두 검토하여 얻은 의사-정답(pseudo-ground-truth) 노출 시간 벡터이다. 이를 구하는 방법은 다음과 같다. 먼저, 미리 정의된 노출 시간 조합 $\hat{\mathbf{t}}_i$ 를 원소로 갖는 집합 $\mathcal{E} = \{\hat{\mathbf{t}}_1, \dots\}$ 를 설정한다.

본 실험에서는 균일·비균일 노출 시간 조합을 모두 고려하여 $\mathcal{E} = \{(8, 8, 8, 8), (12, 12, 12, 12), (16, 16, 16, 16), (24, 24, 24, 24), (32, 32, 32, 32), (8, 12, 18, 27), (8, 16, 32, 64), (12, 18, 27, 41), (12, 24, 48, 96), (16, 24, 36, 54)\}/1920$ 로 정의하였다. 다음으로, $\mathcal{D}_{\text{exposure}}$ 에서 장면 복사조도 시퀀스 \mathbf{S} 를 샘플링하고 이로부터 노출 예측 모듈의 입력으로 사용되는 게인 g_p ,

프리뷰 이미지 I_p 및 I'_p , 모션 크기 m_p , 그리고 정답 이미지 I_{gt} 를 합성한다.

그런 다음, \mathcal{E} 내 각 노출 시간 조합에 대해 버스트 이미지를 합성하고, 버스트 이미지 복원을 수행한 뒤 그 결과를 I_{gt} 와 비교한다. 마지막으로, 각 이미지별로 가장 높은 PSNR을 보인 노출 시간 조합을 해당 이미지에 대한 $\mathbf{t}_{\text{pseudo-gt}}$ 로 설정한다.

이후, 앞서 합성한 m_p 와 g_p 를 이차식에 입력하면 버스트 이미지의 노출 시간 t_1 과 배율 r 이 출력되며 이는 다음과 같은 수식으로 표현할 수 있다:

$$t_1 = a_1 \cdot g_p^2 + a_2 \cdot m_p^2 + a_3 \cdot g_p m_p + a_4 \cdot g_p + a_5 \cdot m_p + a_6 \quad (5)$$

$$r = a_7 \cdot g_p^2 + a_8 \cdot m_p^2 + a_9 \cdot g_p m_p + a_{10} \cdot g_p + a_{11} \cdot m_p + a_{12} \quad (6)$$

여기서 변수 a_i ($i = 1, 2, \dots, 12$)는 평균 제곱 오차(mean squared error)를 통해 선형 회귀로 최적화된다.

4.2 노출에 따른 복원 성능 비교 실험

Table 1. Quantitative results of restoration performance according to the exposure settings of burst images. Here, t_1 denotes the exposure time of the first burst image, and r represents the ratio parameter. The unit of exposure times is 1/1920 seconds.

(t_1, r)	PSNR↑	(t_1, r)	PSNR↑	(t_1, r)	PSNR↑
(8, 1)	33.84	(8, 1.5)	35.13	Best (uniform)	35.48
(12, 1)	34.83	(8, 2)	35.17	Best (non-uniform)	35.56
(16, 1)	35.16	(12, 1.5)	35.32	Best (all)	35.63
(24, 1)	35.09	(12, 2)	34.91	Ours	35.50
(32, 1)	34.65	(16, 1.5)	35.23	-	-

이 실험에서는 최적의 노출 시간을 예측하는 것의 중요성을 보이기 위해 버스트 이미지의 노출 방식에 따른 성능을 평가하고 이를 비교 분석하였다. Table 1에서 ‘Best(uniform)’은 각 시퀀스에 대해 가장 좋은 성능을 보이는 균일 노출 시간 조합만을 채택해 평균 낸 성능이다. 예를 들어, 테스트 데이터셋의 특정 시퀀스를 기준으로 5개의 균일 노출 시간 조합($r = 1$)에 해당하는 버스트 이미지들을 각각 합성한 다음 복원 네트워크에 입력하여 5개의 복원된 이미지를 획득한다. 그런 다음, 복원된 이미지들을 정답 이미지와 비교하여 PSNR을 측정하고 이 중 가장 좋은 성능을 해당 시퀀스의 ‘Best(uniform)’ 성능으로 채택한다. ‘Best(non-uniform)’은 5개의 비균일 노출 시간($r \neq 1$) 조합을 사용한 경우이고 ‘Best(all)’은 10개의 노출 시간 조합을 모두 사용한 경우이다. 사전에 정의된 노출 시간을 고정적으로 사용했을 때, 각 노출 시간 조합마다 성능 차이가 꽤 큰 것을 알 수 있다. 또한, 비균일 노출 시간($r \neq 1$)을 사용하는 것이 균일 노출 시간($r = 1$)을 사용하는 것보다 전반적으로 성능이 더 좋으며 ‘Best’ 성능도 약 0.1dB 앞서는 것을 확인할 수 있다.

본 연구에서 제안하는 노출 예측 모듈을 사용할 경우(Ours), 각각의 균일·비균일 노출 시간 조합을 고정적으로 사용했을 때의 평균 성능(34.93 dB)보다 약 0.57 dB 더 좋으며, 각각의 평가 시퀀스에 대해 직접 가장 좋은 성능을 찾은 ‘Best (all)’의 성과와 약 0.1 dB 밖에 차이가 나지 않는 것을 확인할 수 있다. 여기서 ‘Best’는 실제로는 불가능한 시나리오이며 해당 성능을 사람이 직접 찾은 가상의 상한(pseudo upper bound) 정도로 생각하면 된다. 이를 통해, 비균일 노출 시간을 가정하는 것과 촬영 환경에 대한 정보(모션, 게인 등)를 고려하여 버스트 이미지의 노출 시간을 예측하는 것이 더욱 효과적인 버스트 이미지 복원을 수행하는 데에 도움이 됨을 알 수 있다.

Figure 3은 입력으로 사용되는 합성된 버스트 이미지와 그에 대한 복원 결과를 보여준다. 2열의 결과는 본 연구에서 제안하는 노출 예측 모듈을 사용한 경우이며 3열과 4열은 각각 Table 1의 균일 노출 시간 조합과 비균일 노출 시간 조합 중 가장 복원 성능이 좋았던 조합을 사용한 경우이다. 복원 결과를 보면 고정된 노출 시간을 사용할 경우 상호보완적인 정보를 활용하기 어렵기

나 촬영 환경을 반영할 수 없기 때문에 복원 결과의 품질이 떨어지는 것을 알 수 있다. 반면, 본 연구의 방법(Ours)은 촬영 환경을 반영하여 최적의 노출 시간을 예측하므로 복원 성능이 우수하다.

4.3 노출 예측 모듈 구성 요소 검증 실험

Table 2. Quantitative results of restoration performance according to the inputs of the polynomial regression model.

Inputs	w/o m_p	w/o g_p	Full
PSNR↑	35.43	35.48	35.50

입력의 종류 이 실험에서는 노출 예측 모듈에서 다항 회귀 모델의 입력으로 사용되는 모션 정보 m_p 와 게인 g_p 의 영향을 알아보기 위해 입력에 따른 성능을 평가하고 분석하였다. Table 2를 보면 m_p 와 g_p 를 모두 사용했을 때(3열)의 성능이 가장 좋으며 둘 중 한 가지 정보만을 사용할 경우(1-2열) 성능이 하락한다. 특히, m_p 를 사용하지 않을 경우에 성능이 더욱 하락하는데 이는 더블러링 문제가 디노이징 문제보다 상대적으로 어려워 모션 정보 m_p 의 영향력이 더 크기 때문일 것으로 짐작할 수 있다.

회귀 모델의 차수 노출 예측 모듈에서 다항 회귀 모델의 차수에 따른 영향을 알아보기 위해 식 (5)로 정의된 다항 회귀 모델의 차수를 1부터 4까지 변경하며 성능을 평가하였다. 그 결과, 모든 성능이 약 35.50 dB 정도로 거의 동일했으며 다른 실험에서 노출 예측 모듈 다항 회귀 모델을 2차식으로 선정하여 사용하였다. 이는 입력이 모션과 노이즈의 대략적인 양을 표현하는 단순한 정보이기 때문에 입력과 출력 간의 관계 곧, 패턴을 파악하기에 정보가 충분하지 않기 때문임을 짐작할 수 있다. 이러한 문제는 추가적인 정보를 입력으로 사용하거나 학습시 보다 다양한 노출 시간 후보를 고려하고 다층 퍼셉트론(multi-layer perceptron)으로 다항 회귀 모델을 대체함으로써 해결할 수 있을 것이다.

5 결론

본 연구는 촬영 환경을 반영하여 최적의 비균일 노출 시간을 자동으로 산출하는 새로운 비균일 노출 예측 파이프라인을 제안한다. 본 연구의 노출 예측 모듈은 카메라의 프리뷰로부터 매우 짧은 시간 내에 비균일 노출 시간 조합을 계산하며, 이를 통해 고정 노출 시간 대비 향상된 복원 성능을 달성한다. 본 모듈은 실시간 예측이 가능하고 연산량이 적으므로 다양한 응용 분야(저조도 환경, 모바일 카메라 등)에 효과적으로 활용될 수 있다.

한계점과 미래 연구 방향 제안하는 모듈은 게인과 모션만으로 노출 시간을 추정하므로 입력 환경을 정밀하게 파악하기에 정보가 부족할 수 있다. 이는 추가 정보를 입력으로 사용함으로써

보완이 가능하다. 또한, 다항 회귀 모델은 t_1 과 r 을 독립적으로 계산하기 때문에 두 변수 간 상관관계를 고려하지 못한다. 이러한 문제는 다항 회귀 모델을 다층 퍼셉트론 기반 다변량 예측으로 대체함으로써 개선할 수 있다.

감사의 글

본 연구는 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원 받아 수행된 연구임 (No.2019-0-01906, 인공지능대학원지원(포항공과대학교)). 본 연구는 또한 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터사업의 지원으로 수행되었음 (IITP-2024-00437866).

References

- [1] A. Dudhane, S. W. Zamir, S. Khan, F. S. Khan, and M.-H. Yang, “Burst image restoration and enhancement,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5759–5768.
- [2] —, “Burstformer: Burst image restoration and enhancement transformer,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.
- [3] S. Nah, S. Son, and K. M. Lee, “Recurrent neural networks with intra-frame iterations for video deblurring,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [4] Z. Zhong, Y. Gao, Y. Zheng, B. Zheng, and I. Sato, “Efficient spatio-temporal recurrent neural network for video deblurring,” in *The European Conference on Computer Vision (ECCV)*, 2020.
- [5] M. Delbracio and G. Sapiro, “Burst deblurring: Removing camera shake through fourier burst accumulation,” in *CVPR*, 2015.
- [6] M. Aittala and F. Durand, “Burst image deblurring using permutation invariant convolutional neural networks,” in *ECCV*, 2018.
- [7] G. Bhat, M. Danelljan, L. Van Gool, and R. Timofte, “Deep burst super-resolution,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 9209–9218.
- [8] K. C. Chan, X. Wang, K. Yu, C. Dong, and C. C. Loy, “Basicvsr: The search for essential components in video super-resolution and beyond,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.
- [9] K. C. K. Chan, S. Zhou, X. Xu, and C. C. Loy, “Basicvsr++: Improving video super-resolution with enhanced propagation and alignment,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.
- [10] R. Wu, Z. Zhang, S. Zhang, H. Zhang, and W. Zuo, “Rbsr: Efficient and flexible recurrent network for burst super-resolution,” in *PRCV*, 2023.
- [11] Z. Zhang, S. Zhang, R. Wu, Z. Yan, and W. Zuo, “Exposure bracketing is all you need for a high-quality image,” in *ICLR*, 2025.
- [12] N. K. Kalantari and R. Ramamoorthi, “Deep high dynamic range imaging of dynamic scenes,” *ACM Transactions on Graphics (TOG)*, 2017.
- [13] Q. Yan, D. Gong, Q. Shi, A. van den Hengel, C. Shen, I. Reid, and Y. Zhang, “Attention-guided network for ghost-free high dynamic range imaging,” in *CVPR*, 2019.
- [14] K. R. Prabhakar, R. Arora, A. Swaminathan, K. P. Singh, and R. V. Babu, “A fast, scalable, and reliable deghosting method for extreme exposure fusion,” in *IEEE International Conference on Computational Photography (ICCP)*, 2019.
- [15] S. Wu, Y.-W. Tai, and C.-K. Tang, “Deep high dynamic range imaging with large foreground motions,” in *ECCV*, 2018.
- [16] Y. Niu, J. Wu, Z. Guo, and R. W. H. Lau, “Hdr-gan: Hdr image reconstruction from multi-exposed ldr images with large motions,” *IEEE Transactions on Image Processing (TIP)*, 2021.
- [17] Z. Liu, Y. Wang, and B. Zeng, “Ghost-free high dynamic range imaging with context-aware transformer,” in *ECCV*, 2022.
- [18] Q. Yan, W. Chen, S. Zhang, Y. Zhu, and J. Zhang, “A unified hdr imaging method with pixel and patch level,” in *CVPR*, 2023.
- [19] S. Tel, Z. Wu, Y. Zhang, B. Heyrman, R. Timofte, and D. Ghahic, “Alignment-free hdr deghosting with semantics consistent transformer,” in *ICCV*, 2023.
- [20] Z. Zhang, H. Wang, S. Liu, X. Wang, L. Lei, and W. Zuo, “Self-supervised high dynamic range imaging with multi-exposure images in dynamic scenes,” in *ICLR*, 2024.

- [21] J. W. Song, Y.-I. Park, K. Kong, J. Kwak, and S.-J. Kang, “Selective transh-based selective hdr imaging using ghost region mask,” in *ECCV*, 2022.
- [22] L. Yuan, J. Sun, L. Quan, and H.-Y. Shum, “Image deblurring with blurred/noisy image pairs,” in *SIGGRAPH*, 2007.
- [23] M. Chang, H. Feng, Z. Xu, and Q. Li, “Low-light image restoration with short-and long-exposure raw pairs,” *IEEE Transactions on Multimedia (TMM)*, 2021.
- [24] J. Mustaniemi, J. Kannala, J. Matas, S. Särkka, and J. Heikkilä, “Lsd2–joint denoising and deblurring of short and long exposure images with cnns,” in *BMVC*, 2020.
- [25] Y. Zhao, Y. Xu, Q. Yan, D. Yang, X. Wang, and L.-M. Po, “D2hnet: Joint denoising and deblurring with hierarchical network for robust night image restoration,” in *ECCV*, 2022.
- [26] Z. Zhang, R. Xu, M. Liu, Z. Yan, and W. Zuo, “Self-supervised image restoration with blurry and noisy pairs,” in *NeurIPS*, 2022.
- [27] S. Shekarforoush, A. Walia, M. A. Brubaker, K. G. Derpanis, and A. Levinshtein, “Dual-camera joint deblurring-denoising,” *arXiv preprint arXiv:2309.08826*, 2023.
- [28] W.-S. Lai, Y. Shih, L.-C. Chu, X. Wu, S.-F. Tsai, M. Krainin, D. Sun, and C.-K. Liang, “Face deblurring using dual camera fusion on mobile phones,” *ACM Transactions on Graphics (TOG)*, 2022.
- [29] Z. Teed and J. Deng, “Raft: Recurrent all-pairs field transforms for optical flow,” in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*. Springer, 2020, pp. 402–419.
- [30] S. Nah, T. Hyun Kim, and K. Mu Lee, “Deep multi-scale convolutional neural network for dynamic scene deblurring,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3883–3891.
- [31] J. Rim, G. Kim, J. Kim, J. Lee, S. Lee, and S. Cho, “Realistic blur synthesis for learning image deblurring,” in *European conference on computer vision*. Springer, 2022, pp. 487–503.
- [32] W. Kim, G. Kim, J. Lee, S. Lee, S.-H. Baek, and S. Cho, “Paramisp: Learned forward and inverse isps using camera parameters,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 26 067–26 076.
- [33] T. Brooks, B. Mildenhall, T. Xue, J. Chen, D. Sharlet, and J. T. Barron, “Unprocessing images for learned raw denoising,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 11 036–11 045.
- [34] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, “Automatic differentiation in PyTorch,” in *Proceedings of the Neural Information Processing Systems Workshops (NeurIPSW)*, 2017.
- [35] I. Loshchilov and F. Hutter, “Decoupled weight decay regularization,” in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2019.
- [36] ———, “Sgdr: Stochastic gradient descent with warm restarts,” *arXiv preprint arXiv:1608.03983*, 2016.

〈 저자 소개 〉



이 영 기

- 2022년 ~ 현재 포항공과대학교 컴퓨터공학과 학사
- <https://orcid.org/0009-0000-3128-1990>



김 우 혁

- 2022년 세종대학교 인공지능학과 학사
- 2022년 ~ 현재 포항공과대학교 인공지능대학원 통합과정
- <https://orcid.org/0009-0006-5691-5447>



조 성 현

- 2005년 8월 포항공과대학교 컴퓨터공학 학사
- 2012년 2월 포항공과대학교 컴퓨터공학 박사
- 2012년 3월 ~ 2014년 3월 Adobe Research 연구원
- 2014년 4월 ~ 2017년 4월 삼성전자 책임연구원
- 2017년 4월 ~ 2019년 8월 대구경북과학기술원 조교수
- 2019년 8월 ~ 2021년 8월 포항공과대학교 조교수
- 2021년 9월 ~ 현재 포항공과대학교 부교수
- <https://orcid.org/0000-0001-7627-3513>