한국컴퓨터그래픽스학회
Korea Computer Graphics Society Vol. 32, No. 3, P. 155~162

신경망 기반 디블러링을 이용한 선명한 동적 복사장 재구성

권강민° 양희민 조성현*

포항공과대학교

{gmkwon99, heeminid, s.cho}@postech.ac.kr

Sharp Dynamic Radiance Field Reconstruction via Learning-based Video Deblurring

Gangmin Kwon ° Heemin Yang Sunghyun Cho*

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

요 약

블러가 있는 단안 비디오로부터 선명한 동적 복사장을 재구성하는 것은 novel view synthesis (NVS)에서 중요한 문제이다. 최근 연구들은 동적 복사장과 블러 발생 원인을 함께 최적화하는 접근 방식을 주로 사용한다. 그러나 이러한 방식은 별도의 정보 없이 블러의 원인과 복사장을 동시에 추정해야 하므로, 심한 블러가 존재하는 경우 안정적인 복원이 어렵다. 본 논문에서는 이러한 문제를 완화하기 위해 학습 기반 비디오 디블러링과 동적 복사장 최적화를 순차적으로 진행하는 파이프라인을 제안한다. 먼저 디블러링 모델을 이용하여 블러가 포함된 입력 비디오를 복원하고, 복원된 프레임으로부터 초기 동적 복사장을 재구성한다. 이후 초기 동적 복사장에서 렌더링한 이미지를 복사장 가이드 기반 비디오 디블러링 모델의 가이드로 활용하여 입력 비디오를 한 번 더 복원한다. 최종적으로, 복원된 프레임을 이용하여 최종 동적 복사장을 최적화한다. 실험을 통해 제안 방법은 새로운 시점의 합성 품질을 PSNR 및 LPIPS 측면에서 향상시키며, 심한 모션 블러가 존재하는 영역에서도 시각적으로 더 선명한 결과를 생성함을 보인다.

Abstract

Reconstructing a sharp dynamic radiance field from blurry monocular video is an important problem in novel view synthesis (NVS). Recent approaches jointly optimize dynamic scene representations with blur formation models. However, these methods often suffer from unstable optimization and struggle to recover sharp results under severe blur. To alleviate this problem, we propose a sequential pipeline that combines neural network-based video deblurring with dynamic radiance field optimization. Our method first applies initial deblurring to the input video and learns a dynamic radiance field from the deblurred frames. It then renders images from the learned radiance field and uses them as guidance for a RF-guided video deblurring model to restore the input video again. Finally, the restored sharp frames are used to optimize the final dynamic radiance field. Experimental results show that the proposed method achieves higher PSNR and lower LPIPS than existing methods in novel view synthesis, while demonstrating robust restoration performance in regions with severe motion blur.

키워드: 디블러링, 복사장, 4DGS
Keywords: deblurring, novel view synthesis, 4DGS

1 서론

Novel view synthesis (NVS)는 입력된 이미지들로부터 새로운 시점의 이미지를 합성하는 문제로, 증강현실이나 가상현실, 3D 콘텐츠 등 다양한 시나리오에서 활용될 수 있다. Neural Radiance Field (NeRF) [1] 와 3D Gaussian Splatting [2]와 같이 이미지들로부터 복사장을 최적화하여 다른 뷰의 이미지를 만드는 방법들이 높은 성능을 보여주면서, 이 방법들을 기반으로 다양한 복사장 최적화 연구들이 진행되어 왔다. 그러나 이러한 연구들은 입력

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

Received : 2026.05.29./ Review completed : 1st 2026.06.07. 2nd 2026.06.25./ Accepted : 2026.06.29.
DOI : 10.15701/kcgs.2026.32.3.155
ISSN : 1975-7883(Print)/2383-529X(Online)

이미지에 열화가 없이 선명하게 촬영되었다고 가정하며, 이미지에 손상이 있을 경우 전체 복사장의 품질 저하로 이어진다. 실제 촬영 환경에서는 이미지의 품질을 저하하는 다양한 요소가 존재하며, 특히 카메라의 흔들림이나 렌즈의 초점 문제로 인하여 블러가 많이 발생한다.

블러가 있는 입력에 대하여 선명한 복사장을 만드는 복사장 디블러링 문제를 해결하기 위하여, 블러를 최적화 가능한 요소들로 모델링하여 복사장과 함께 최적화하는 기법들이 제안되었다. 일부 연구는 블러가 커널에 의해 형성되었다고 가정하고 복사장과 커널을 동시에 최적화한다 [3, 4, 5]. 또 다른 연구들은 블러가 카메라의 움직임에 의해 형성되었다고 가정하여, 촬영 시간동안의 가상 카메라의 위치를 복사장과 함께 최적화한다 [6, 7, 8, 9, 10].

최근 동적 복사장에 대한 연구들이 [11, 12, 13] 등장하며, 블러가 있는 입력에서 동적 복사장을 최적화하는 연구들이 제안되고 있다 [14, 15, 16, 17, 18, 19]. 이들은 블러가 있는 단안 비디오를 입력으로 받아, 선명한 새로운 시점의 이미지를 만들 수 있는 동적 복사장을 최적화하는 것을 목적으로 한다. 이들은 기존의 최적화 방식들에 동적 복사장의 최적화를 결합하여 블러 모델링과 복사장, 시간에 대한 복사장의 변화를 한 번에 최적화한다. 이러한 연구들은 우수한 성능을 보여주지만, 최적화하는 항들이 많아져 최적화 과정이 복잡해지고, 디블러링에 실패하거나 원치 않는 아티팩트가 발생할 수 있다.

한편, 이미지/비디오 디블러링 분야는 신경망과 결합되며 계속해서 발전하고 있다 [20, 21, 22, 23, 24, 25]. 이들은 대규모 데이터셋으로부터 data-driven prior을 학습하기에 복잡한 형태의 블러에도 효과적으로 대응할 수 있다. 특히, 비디오 디블러링 방법들은 인접한 프레임들이 유사한 장면 구조와 정보를 공유하기 때문에, 서로 정보를 교환하며 단일 이미지 기반 방법보다도 더 잘 복원한다. 그러나, 일반적인 디블러링 기법들은 주로 PSNR 및 SSIM과 같은 픽셀 단위 오차 기반 지표에 초점을 맞추며, 복사장 최적화에 필요한 3D 공간상의 대응 관계 등에 주목하지 않는다. 그래서 디블러링 기법을 직접적으로 복사장의 최적화에 적용할 경우 3D 공간상에서 프레임 간의 불일치가 발생한다고 알려져 있다. DDRF [26]는 최적화된 복사장의 결과를 다시 신경망의 입력으로 삼는 재귀적인 프레임워크를 통해 프레임 간의 불일치 문제를 보완하였다. 그러나, DDRF는 움직이는 물체가 없는 상황을 가정하기에 블러가 있는 단안 비디오에서 동적 복사장을 최적화하는 문제에 적용하기에는 어려움이 있다.

이러한 문제를 해결하기 위해, 본 논문에서는 모션 블러가 존재하는 단안 비디오로부터 고품질의 4D Gaussian Splatting (4DGS)을 학습하기 위한 통합 프레임워크를 제안한다. 우리의 핵심 아이디어는 동적 복사장 최적화와 카메라와 객체의 움직임에 의한 블러가 결합된 모호한 문제를 data-driven prior라는 추가적인 정보를 가진 디블러링과 동적 복사장 최적화 문제로 분리하는 것이다. 또한, 디블러링 모델이 4DGS의 렌더 정보를 추가로 받아 3D 공간상의 일관성을 이해할 수 있는 비디오 디블러링 모델인 복사장 가이드 기반 비디오 디블러링 모델(RF-guided video deblurring model)을 제안한다. 제안하는 디블러링 모델은 단순히 각 프레임을 독립적으로 선명하게 만드는 것이 아니라, 4D 장면 표현으로부터 얻은 구조적 정보를 함께 사용하여 동적 NVS 학습에 더 적합하고 기하학적으로 일관된 디블러링 결과를 생성한다.

본 논문의 기여는 다음과 같다.

- 우리는 모션 블러가 있는 단안 비디오에서, 학습된 신경망 기반 디블러링 모듈과 4DGS을 결합한 순차적 파이프라인의 가능성을 보인다.
- 우리는 4DGS 렌더링 결과를 보조 정보로 활용하여 비디오 디블러링 결과를 개선하는 복사장 가이드 기반 비디오 디블러링 모델을 제안한다.
- 블러가 있는 단안 비디오 데이터셋에서 해당 방법은 우수한 정량적, 정성적 성능을 보인다.

2 관련 연구

2.1 복사장 디블러링

DeblurNeRF [3]에서 NeRF와 블러를 동시에 최적화하는 방향을 제안한 이후, 복사장 디블러링 방식에 대한 많은 연구가 진행되었다. 블러를 커널의 형태로 모델링하여 최적화하는 기법들 [3, 4, 5]은 카메라 모션 블러와 디포커스 블러를 각각 모델링하며, 두 블러에 모두 적용한다. 한편, 블러를 노출 시간동안의 카메라 움직임으로 모델링하는 기법들 [6, 7, 8, 9, 10] 역시 존재한다. 이러한 방식들은 블러가 있는 단안 비디오를 입력으로 받는 동적 복사장 디블러링 방식으로 확장되었다 [16, 17, 18, 19]. Zhang et al. [16]은 Feature Extractor과 BP-Net을 이용하여 블러 커널을 근사하고, 이를 동적 복사장과 동시에 최적화한다. BARD-GS [17]는 동적 장면에서 발생하는 모션 블러를 카메라 모션에 의한 블러와 객체 모션에 의한 블러로 분리하여 모델링한다. Deblur4DGS [18]는 노출 시간 동안의 연속적인 장면 변화를 모델링하고, 여러 해상도를 이용하여 동적 복사장을 복원한다. MoBGS [19]는 잠재 카메라 궤적과 노출 인식 모션을 추정하여 전역 카메라 모션 블러와 지역 객체 모션 블러를 함께 보정한다. 이러한 방식들은 동시에 많은 것들을 학습해야 하기에 많은 계산 시간을 요구하며, 특히 시간에 따라 움직이는 물체의 경우 타 프레임에서 적절한 대응 관계를 얻기 힘들어 더욱 제한적이다.

한편, DDRF [26]는 위 논문들과 달리 이미지 디블러링 네트워크를 사용하여 디블러링과 복사장의 최적화를 반복적으로 진행하는 프레임워크를 제안하였다. 이러한 방식은 data-driven prior을 추가적으로 활용함과 동시에, 최적화할 항들이 적어 보다 빠른 시간에 동작한다. 그러나 이는 움직이는 물체가 없고 입력된 이미지들이 연속적이지 않은 상황을 가정하며, 블러가 있는 단안 비디오를 입력으로 받는 상황을 가정하지는 않는다.

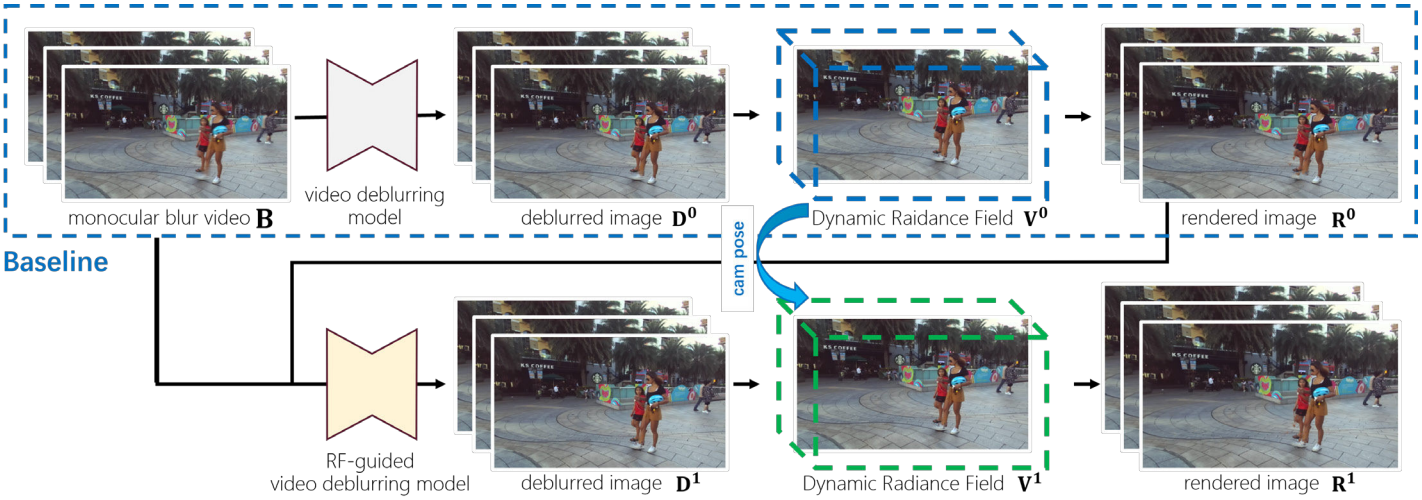

Figure 1. Overall pipeline structure of the proposed method.

2.2 신경망 기반 디블러링

이미지의 품질을 저하시키는 블러를 해결하기 위한 디블러링 연구는 과거부터 계속 연구되어 왔다. 최근에는 대규모 데이터셋을 이용하여 딥러닝 모델을 학습하는, data-driven prior을 이용하여 디블러링하는 접근법이 주류를 이룬다 [20, 21]. 한편, 비디오 디블러링은 이미지 디블러링 방법론에 더하여, 입력 이미지가 짧은 시간 간격, 비슷한 카메라 위치에서 학습되었다는 가정을 통해 주변 이미지로부터 상호 보완되는 정보를 활용한다 [22, 23, 24, 25]. 기존의 정적 복사장 연구의 다중 뷰 세팅은 카메라 포즈가 비디오에 비해 크게 변화하여 사용하기 어려웠으나, 단안 비디오를 입력으로 주로 받는 동적 복사장 세팅에서는 매우 유용하게 활용될 수 있다.

3 방법

3.1 사전 지식

메소드 설명에 앞서, 모델의 기반으로 사용되는 최신 비디오 디블러링 네트워크인 BSSTNet [25] 을 먼저 소개한다. BSSTNet은 연속된 프레임 간의 특징 벡터를 상호 전파하여 이미지를 선명하게 만드는 Blur-aware Bidirectional Feature Propagation (BBFP) 모듈과, 트랜스포머를 이용하여 이미지를 선명하게 만드는 Blur-aware Spatio-temporal Sparse Transformer (BSST) 모듈로 구성된다. BBFP 모듈은 주변 프레임들로부터 특징 벡터들을 전달받아, 정합하고 개선하는 Blur Feature Alignment (BFA) 모듈로 구성되어 있다. BFA 모듈이 시작 프레임부터 끝 프레임까지 특징 벡터를 전파하는 단계를 하나의 브랜치라 부르며, BBFP는 정방향/역방향으로 교차되어 진행되는 다수의 브랜치들로 구성되어 있다. 블러 이미지 $\mathbf{B}$는 인코더를 통과한 뒤 이미지 특징 벡터의 형태가 되어 BBFP 모듈에서 개선된다. BBFP 모듈의 j번째 브랜치, 시간 t의 특징 벡터를 $\hat{F}_t^j$라 정의하자. BFA는 $\hat{F}_t^j$를 추정하기 위하여 이전 단계의 시간 t 특징 벡터인 $\hat{F}_t^{j-1}$, 현재 브랜치의 이전 시간 특징 벡터들 $\hat{F}_{t-1}^j$, $\hat{F}_{t-2}^j$ 를 이용한다. 세부 모델에는 각 특징 벡터를 다른 시간의 특징 벡터로 매핑하기 위한 추가적인 요소들이 있으나, 편의상 특징 벡터들만으로 모듈을 단순하게 표현하면 다음과 같다.

$$\hat{F}_t^j = \text{BFA}(\hat{F}_t^{j-1}, \hat{F}_{t-1}^j, \hat{F}_{t-2}^j, ...) \tag{1}$$

BSST 모듈은 BBFP 단계를 거치며 정제된 특징 벡터를 트랜스포머 [27]을 이용하여 추가로 정제하는 단계이다. 입력된 이미지 영역을 타일로 나눈 뒤, 블러가 있는 영역을 쿼리로, 선명한 영역을 키/밸류로 삼아 어텐션을 진행하여, 블러가 있는 영역에 선명한 영역의 정보를 주입한다.

3.2 전체 프레임워크

이 연구는 블러가 있는 단안 비디오를 입력으로 받아, 선명한 새로운 시점 이미지를 합성하거나, 그러한 이미지를 생성할 수 있는 선명한 동적 복사장을 추정하는 문제를 다룬다. Figure 1 은 파이프라인의 전반적인 구조를 보여준다. 먼저 M장의 블러가 있는 단안 비디오 프레임들 $\mathbf{B} = \{B_1, \ldots, B_M\}$를 입력으로 받아, 비디오 디블러링을 진행하여 디블러된 비디오 프레임 $\mathbf{D^0} = \{D_1^0, ..., D_M^0\}$를 추정한다 (섹션 3.3). 이후 추정된 디블러된 비디오 프레임 $\mathbf{D^0}$를 입력으로 하여 선명한 복사장 $\mathbf{V}^0$를 학습한다 (섹션 3.4). 최적화가 완료된 $\mathbf{V}^0$로부터 $\mathbf{D}^0$와 똑같은 시점에서 이미지들을 렌더링한다. 그 후, 복사장 가이드 기반 비디오 디블러링 모델은 렌더링된 이미지들을 디블러링 과정에서의 추가적인 보조로 활용하여 $\mathbf{B}$를 디블러링한다 (섹션 3.5). 복사장 가이드 기반 비디오 디블러링 모델을 통해 디블러링이 완료된 이미지들을 가지고 최종 복사장인 $\mathbf{V}^1$을 최적화한다 (섹션 3.4).

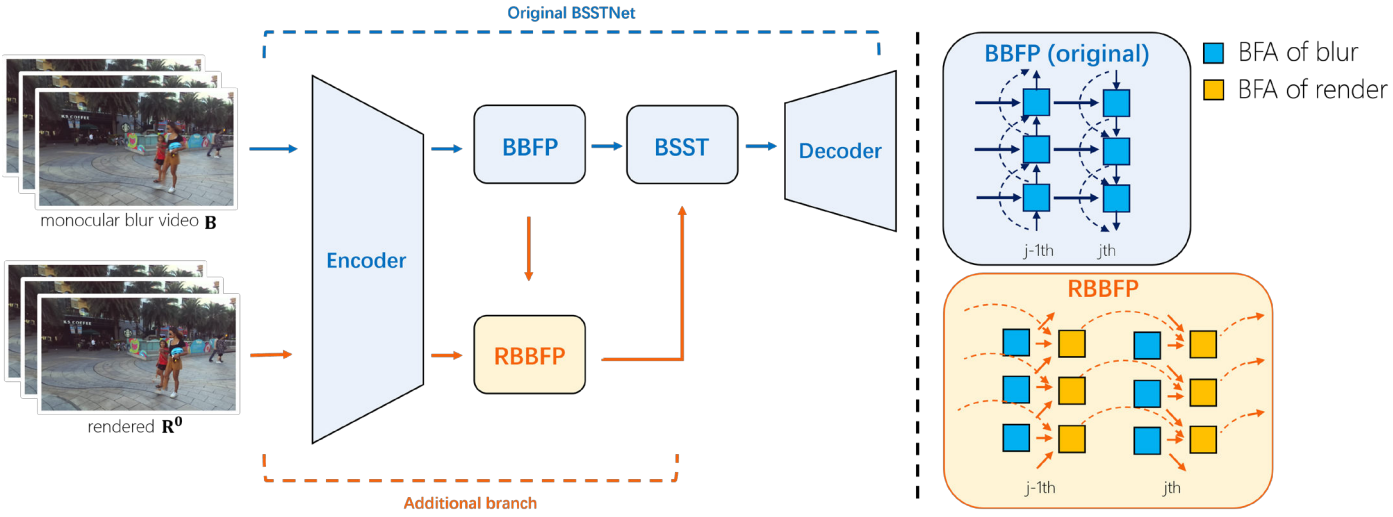

Figure 2. Structure of RF-guided video deblurring model. The blue modules indicate the original components of BSSTNet, while the yellow modules denote the added RBBFP module and render feature flow.

3.3 초기 비디오 디블러링

초기 비디오 디블러링 단계는 블러가 있는 입력 비디오 $\mathbf{B}$에 대하여, 비디오 디블러링 모델을 이용하여 블러를 제거하여, 디블러된 이미지 $\mathbf{D^0} = \{D_1^0, ...D_M^0\}$을 얻는 단계이다. 이는 이후 학습될 복사장 $\mathbf{V}^0$을 덜 블러하게 만드는 데 도움을 준다. 우리는 비디오 디블러링 네트워크인 BSSTNet [25] 을 재학습하여 진행하였다.

3.4 동적 복사장 최적화

동적 복사장 최적화 단계는 각 디블러링 모델로부터 얻은 디블러링 결과 이미지들을 이용하여 동적 복사장을 학습하는 단계이다. $\mathbf{D}^0$를 이용하여 첫 번째 복사장 $\mathbf{V}^0$를 학습하고, $\mathbf{D}^1$를 이용하여 두 번째 복사장 $\mathbf{V}^1$을 학습한다. 각 단계 모두 기존의 4DGS 모델인 Mosca [13] 를 별도의 변경 없이 사용한다.

첫 번째 복사장은 초기 비디오 디블러링 과정을 통해 얻은 $\mathbf{D}^0$를 Mosca에 입력하여 초기 동적 복사장 $\mathbf{V}^0$를 학습한다. 이 단계의 입력인 $\mathbf{D}^0$는 비디오 디블러링 과정을 통해 블러가 완화된 상태지만, 각 이미지 간 일관성이 부족하여 학습되는 전체 기하 구조가 흔들린다. 이 때문에 $\mathbf{V}^0$ 를 바로 최종 결과로 사용하는 대신, 다음 단계 디블러링의 참조 이미지를 생성하기 위한 중간 복사장으로 사용된다. 이를 위하여 입력 비디오와 동일한 시점의 이미지 $\mathbf{R}^0$을 렌더링한다. 이 과정에서는 별도의 카메라 포즈를 제공하지 않고, Mosca가 카메라 포즈와 동적 복사장을 함께 최적화하도록 한다.

두 번째 단계에서도 복사장 가이드 기반 비디오 디블러링 모델의 결과인 $\mathbf{D}^1$을 이용하여 Mosca를 학습시켜 최종 동적 복사장 $\mathbf{V}^1$을 얻는다. 복사장 가이드 기반 비디오 디블러링 모델은 기존 디블러 모델의 일관성 문제를 완화하기에, 이를 유사 pseudo-sharp로 가정하고 최종 복사장 $\mathbf{V}^1$을 학습한다. 이 단계에서는 이전 단계의 Mosca가 학습한 카메라 포즈로 카메라 포즈를 초기화한다.

3.5 복사장 가이드 기반 비디오 디블러링 모델

비디오 디블러링 모델은 비디오의 블러를 완화하지만, 복사장 최적화에 필요한 3D 기하 구조와 다중 시점 간의 일관성을 명시적으로 고려하지 않는다. 복사장 가이드 기반 비디오 디블러링 모델은 3D 상에서 학습되는 이전 단계의 동적 복사장의 정보를 활용하여 이러한 문제를 완화하는 모델이다. 구체적으로 복사장 가이드 기반 비디오 디블러링 모델은 초기 동적 복사장 $\mathbf{V}^0$의 렌더된 이미지 $\mathbf{R}^0$을 기존의 블러 이미지 $\mathbf{B}$와 함께 입력으로 받아 개선된 디블러 이미지 $\mathbf{D^1}$ 을 출력한다.

렌더된 이미지 $\mathbf{R}^0$는 모델 내부에서 기하적으로 일관된 pseudo-sharp 이미지의 역할을 맡는다. 선명한 이미지끼리 매칭할 경우 각 점을 일대일로 매칭할 수 있지만, 블러는 한 점이 여러 픽셀에 도달한 결과이기에 블러가 있는 이미지끼리 매칭할 경우 다대다 형태로 매칭되어야 한다. 복사장 가이드 기반 비디오 디블러링 모델은 $\mathbf{R}^0$를 pseudo-sharp로 이용하여, 비교적 선명한 렌더된 이미지를 기준으로 정합되도록 한다.

Figure 2 는 모델의 전체 구조를 보여준다. 이를 위해, 기존 비디오 디블러 모델인 BSSTNet에 $\mathbf{R^0}$을 사용하는 모듈인 Render-BBFP (RBBFP)를 추가하였다.

RBBFP 모듈은 기존 BSSTNet의 BBFP모듈과 동일하게 여러 BFA 모듈의 브랜치로 구성되어 있다. 블러 이미지의 특징 벡터와 렌더 이미지의 특징 벡터를 구분하기 위해, RBBFP 모듈의 j번째 브랜치에서, 시간 t의 특징 벡터를 $\hat{f}_t^j$이라 하자. RBBFP 모듈의 BFA는 $\hat{f}_t^j$를 추정하기 위하여 이전 브랜치의 렌더링 이미지 특징 벡터인 $\hat{f}_t^{j-1}$과, BBFP 모듈에서 만들어진 동일한 브랜치의 특징 벡터인 $\hat{F}_t^j$, 그리고 그 이전 시간의 특징 벡터인 $\hat{F}_{t-1}^j$ 를 이용한

Table 1: Dynamic deblurring novel view synthesis evaluation of Stereo Blur Dataset

Model	PSNR ↑	LPIPS ↓
Deblur4DGS	27.84	0.066
MoBGS	28.80	0.050
Baseline	28.03	0.048
Ours w/o RF-guided	28.87	0.040
Ours	29.04	0.034

다. 이를 앞의 BFA처럼 표현하면 아래와 같이 표현할 수 있다.

$$\hat{f}_t^j = \text{BFA}(\hat{f}_t^{j-1}, \hat{F}_t^j, \hat{F}_{t-1}^j, ...) \quad (2)$$

4 실험

4.1 실험 상세

테스트 데이터셋 우리는 테스트 데이터셋으로 기존의 동적 복사장 디블러링 논문들 [18, 19] 에서 공통적으로 사용되는 Stereo Blur Dataset [14]을 활용하였다. 이 데이터셋은 ZED 스테레오 카메라로 60fps로 촬영된 데이터셋으로, 좌측 카메라는 블러 생성을 위해, 우측 카메라는 테스트를 위한 새로운 시점으로 이용한다. 좌측 카메라의 이미지를 블러하게 합성하기 위하여, 비디오 프레임 보간 [28] 을 이용하여 480fps로 증가시키고, 17, 33, 49장 중 랜덤하게 프레임 수를 정하여 블러를 합성한 데이터셋이다.

학습 데이터셋 우리는 학습용 데이터셋으로 테스트 데이터셋과 동일한 형태로 구성된 DAVNET [29]의 데이터셋을 사용하였다. 테스트에 사용되는 Stereo Blur Dataset은 학습에 사용하는 DAVNET 데이터셋의 일부이기 때문에 공정한 비교를 위하여 테스트 데이터셋과 중복 혹은 유사한 장면들을 학습용 데이터셋에서 모두 배제하였다. 최종적으로, 65개의 scene을 학습에 사용하였으며, 원본 이미지 크기 (1280 x 720)과 테스트 이미지 크기 (512 x 288)으로 조정한 이미지를 둘 다 학습에 사용하였다.

학습 세부 사항 전체 파이프라인에서 첫 번째 복사장은 이후 디블러링 모델에서 사용할 가이드를 출력하는 모델로, 최종 모델보다는 기하학적 정합에 초점을 둔다. 이 때문에, 첫 번째 복사장은 두 번째 복사장에 비해 가볍고 빠르게 학습된다. 구체적으로, 첫 번째 복사장은 가우시안 스플래팅의 조화 진동수를 1로 설정하여 학습되며, 별도의 카메라 포즈 입력을 받지 않고 자체적으로 카메라 위치를 추정한다. 반면, 두 번째 복사장은 조화 진동수를 3으로 하고, 이전 단계로의 복사장이 최적화한 카메라 포즈를 이용하여 카메라 포즈를 초기화한다.

Table 2: Runtime and rendering speed comparison.

Model	deblurring(m)	Dynamic Radiance Field(m)	Rendering (fps)
MoBGS	X	90	240
Ours(initial)	1	30	50
Ours(RF-guide)	1	50	50

4.2 기존 방법들과의 비교

우리는 우리의 모델을 기존 방법들 [18, 19]과 비교한다. 또한, 복사장 디블러링 문제를 해결하는 가장 단순한 아이디어인 비디오 디블러링 모델 BSSTNet과 동적 복사장 모델 Mosca를 순차적으로 적용한 파이프라인을 Baseline으로 삼아 비교한다. 이 때, BSSTNet은 우리의 모델과 동일하게 학습된 모델을 사용한다. 이미지 평가 메트릭으로는 PSNR과 LPIPS를 사용한다.

Table 1은 Stereo Blur Dataset [14]의 새로운 시점에서 합성된 이미지에 대한 정량적 평가를 보여준다. 학습된 디블러 모델을 사용한 Baseline도 다른 복사장 디블러링 방법들에 비교할 만한 성능이 나왔으며, 복사장 가이드 기반 비디오 디블러링 모델까지 전체 파이프라인을 진행할 경우 다른 방법보다 높은 성능을 달성한다. Figure 3은 이러한 다른 시점 이미지를 렌더링한 결과 중 일부를 보여준다. 첫 번째 행에서는 다리가 빠르게 움직여 블러가 크게 형성되었을 때, data-driven prior을 사용하는 Ours가 다른 모델들에 비해 더 선명하게 복원하는 것을 보여준다. 두 번째 행의 팔 부분과, 세 번째 행의 다리 부분 주변에서 Deblur4DGS와 MoBGS가 보다 블러한 모습을 보여준다. 반면, 디블러 모델을 사용한 경우 그러한 부분이 적게 드러난다.

4.3 절제 실험 (Ablation Study)

복사장 가이드 기반 비디오 디블러링 모델의 성능을 분석하기 위하여 절제 실험을 진행하였다. Table 1의 Ours w/o RF-guided 항은 전체 파이프라인에서 복사장 가이드 기반 비디오 디블러링 모델을 pretrained BSSTNet으로 대체한 결과이다. Ours는 Ours w/o RF-guided 대비 PSNR을 추가로 향상시키고 LPIPS를 감소시킨다. 이는 복사장 가이드 기반 비디오 디블러링 모델이 렌더링 이미지를 참조로 활용하여 디블러링 결과를 개선하고, 최종 복사장을 학습하는 데 더 적합한 입력을 생성함을 보여준다.

4.4 계산 시간 분석

우리의 파이프라인은 신경망 기반 디블러링 단계를 추가로 포함하므로, 이러한 단계가 전체 수행 시간에 큰 부담을 주는지 확인할 필요가 있다. Table 2는 RTX3090 에서 측정한 각 단계별 수행 시간을 보여준다. 우리의 파이프라인에서 디블러링 단계는 전체 시간 중 작은 비중을 차지하며, 대부분의 계산 비용은 동적 복사장 최적화 과정에서 발생한다.

우리의 파이프라인의 전체 수행시간은 약 80분 정도로, MoBGS의 최적화 시간(90분)과 유사하다. MoBGS와 같이 복사

Figure 3. Visual comparisons for dynamic deblurring novel view synthesis on the Stereo Blur Dataset.

장과 블러 모델링을 동시에 최적화하는 기법은 최적화의 소요 시간이 길다. 반면, 우리의 파이프라인은 디블러링과 복사장 학습을 교차하여 수행하므로 각 단계의 수행 시간은 길지 않지만, 동적 복사장 학습을 여러 번 하여 전체 파이프라인의 소요 시간이 증가하였다. 학습 세부 사항에서 언급한 바와 같이, 첫 번째 복사장은 최종 복사장을 학습하는 두 번째 복사장에 비하여 가볍고 빠르게 학습되었다.

한편, 렌더링 속도 측면에서는 MoBGS가 제안한 방법보다 약 5배 빠른 결과를 보였다. 이러한 차이는 MoBGS의 백본 모델인 SplineGS [12]가 우리의 백본 모델인 Mosca [13]보다 빠른 렌더링 속도를 갖기 때문으로 추정된다. 실제로 SplineGS [12]는 NVIDIA 데이터셋에서 약 400 fps의 렌더링 속도를 보고하였으며, 이는 동일 논문에서 실험한 Mosca의 렌더링 속도 40 fps 대비 약 10배 빠른 수치이다.

5 결론

본 연구에서는 블러를 복사장과 동시에 최적화하는 다른 복사장 디블러링 방법들과 달리, 신경망 기반 비디오 디블러링과 선명한 복사장 최적화를 순차적으로 진행하는 프레임워크를 제안한다. 우리는 복사장 학습에 도움이 되는 디블러링 모델인 복사장 가이드 기반 비디오 디블러링 모델을 제안하였다. 이러한 파이프라인은 테스트 데이터셋에서 우수한 성능을 보이며, 복사장 디블러링 문제에서 비주류에 있던 신경망 기반 디블러링 방법의 잠재성을 보여준다.

5.1 한계

이 파이프라인의 한계는 신경망 기반 디블러링 방법이 학습 데이터셋에 크게 의존한다는 것이다. 이 때문에, 학습 데이터셋에 없던 블러가 등장할 경우 우리의 네트워크는 성능이 하락할 수 있다.

감사의 글

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단(NRF) (RS-2026-25492695), 정보통신기획평가원(IITP) (RS-2026-25517417, 자유시점 미디어 압축·복원·렌더링 기술개발; RS-2019-II191906, 인공지능대학원지원(포항공과대학교))의 지원을 받아 수행된 연구임.

References

[1] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.

[2] B. Kerbl, G. Kopanas, T. Leimkühler, G. Drettakis, *et al.*, "3d gaussian splatting for real-time radiance field rendering." *ACM Trans. Graph.*, vol. 42, no. 4, pp. 139–1, 2023.

[3] L. Ma, X. Li, J. Liao, Q. Zhang, X. Wang, J. Wang, and P. V. Sander, "Deblur-nerf: Neural radiance fields from blurry images," in *Proceedings of the IEEE/CVF conference on*

computer vision and pattern recognition, 2022, pp. 12 861–12 870.

[4] D. Lee, M. Lee, C. Shin, and S. Lee, "Dp-nerf: Deblurred neural radiance field with physical scene priors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 12 386–12 396.

[5] C. Peng, Y. Tang, Y. Zhou, N. Wang, X. Liu, D. Li, and R. Chellappa, "Bags: Blur agnostic gaussian splatting through multi-scale kernel modeling," in *European Conference on Computer Vision*. Springer, 2024, pp. 293–310.

[6] P. Wang, L. Zhao, R. Ma, and P. Liu, "Bad-nerf: Bundle adjusted deblur neural radiance fields," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 4170–4179.

[7] D. Lee, J. Oh, J. Rim, S. Cho, and K. M. Lee, "Exblurf: Efficient radiance fields for extreme motion blurred images," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 17 639–17 648.

[8] L. Zhao, P. Wang, and P. Liu, "Bad-gaussians: Bundle adjusted deblur gaussian splatting," in *European Conference on Computer Vision*. Springer, 2024, pp. 233–250.

[9] W. Chen and L. Liu, "Deblur-gs: 3d gaussian splatting from camera motion blurred images," *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, vol. 7, no. 1, pp. 1–15, 2024.

[10] H. Jang, D. Choi, D. Kim, W. Kang, and M. H. Kim, "Splat-based 3d scene reconstruction with extreme motion-blur," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025, pp. 26 425–26 434.

[11] Q. Wang, V. Ye, H. Gao, W. Zeng, J. Austin, Z. Li, and A. Kanazawa, "Shape of motion: 4d reconstruction from a single video," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025, pp. 9660–9672.

[12] J. Park, M.-Q. V. Bui, J. L. G. Bello, J. Moon, J. Oh, and M. Kim, "Splinegs: Robust motion-adaptive spline for real-time dynamic 3d gaussians from monocular video," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 26 866–26 875.

[13] J. Lei, Y. Weng, A. W. Harley, L. Guibas, and K. Daniilidis, "Mosca: Dynamic gaussian fusion from casual videos via 4d motion scaffolds," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 6165–6177.

[14] H. Sun, X. Li, L. Shen, X. Ye, K. Xian, and Z. Cao, "Dyblurf: Dynamic neural radiance fields from blurry monocular video," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 7517–7527.

[15] M.-Q. V. Bui, J. Park, J. Oh, and M. Kim, "Moblurf: Motion deblurring neural radiance fields for blurry monocular video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

[16] X. Zhang, J. Xiao, and Q. Zhang, "Dynamic gaussian splatting from defocused and motion-blurred monocular videos," *Advances in Neural Information Processing Systems*, vol. 38, pp. 71 096–71 121, 2026.

[17] Y. Lu, Y. Zhou, D. Liu, T. Liang, and Y. Yin, "Bard-gs: Blur-aware reconstruction of dynamic scenes via gaussian splatting," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 16 532–16 542.

[18] R. Wu, Z. Zhang, M. Chen, Z. Yan, and W. Zuo, "Deblur4dgs: 4d gaussian splatting from blurry monocular video," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 40, no. 13, 2026, pp. 10 727–10 735.

[19] M.-Q. V. Bui, J. Park, J. L. Gonzalez, J. Moon, J. Oh, and M. Kim, "Mobgs: Motion deblurring dynamic 3d gaussian splatting for blurry monocular video," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 40, no. 4, 2026, pp. 2480–2489.

[20] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 5728–5739.

[21] L. Chen, X. Chu, X. Zhang, and J. Sun, "Simple baselines for image restoration," in *European conference on computer vision*. Springer, 2022, pp. 17–33.

[22] J. Liang, J. Cao, Y. Fan, K. Zhang, R. Ranjan, Y. Li, R. Timofte, and L. Van Gool, "Vrt: A video restoration transformer," *IEEE Transactions on Image Processing*, vol. 33, pp. 2171–2182, 2024.

[23] J. Liang, Y. Fan, X. Xiang, R. Ranjan, E. Ilg, S. Green, J. Cao, K. Zhang, R. Timofte, and L. V. Gool, "Recurrent video restoration transformer with guided deformable attention," *Advances in Neural Information Processing Systems*, vol. 35, pp. 378–393, 2022.

[24] D. Li, X. Shi, Y. Zhang, K. C. Cheung, S. See, X. Wang, H. Qin, and H. Li, "A simple baseline for video restoration with grouped spatial-temporal shift," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 9822–9832.

[25] H. Zhang, H. Xie, and H. Yao, "Blur-aware spatio-temporal sparse transformer for video deblurring," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 2673–2681.

[26] H. Choi, H. Yang, J. Han, and S. Cho, "Exploiting deblurring networks for radiance fields," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 6012–6021.

[27] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[28] S. Niklaus, L. Mai, and F. Liu, "Video frame interpolation via adaptive separable convolution," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 261–270.

[29] S. Zhou, J. Zhang, W. Zuo, H. Xie, J. Pan, and J. S. Ren, "Davanet: Stereo deblurring with view aggregation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10 996–11 005.

[30] B. Lee, H. Lee, X. Sun, U. Ali, and E. Park, "Deblurring 3d gaussian splatting," in *European Conference on Computer Vision*. Springer, 2024, pp. 127–143.

[31] 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.

[32] S. Su, M. Delbracio, J. Wang, G. Sapiro, W. Heidrich, and O. Wang, "Deep video deblurring for hand-held cameras," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1279–1288.

〈저자소개〉

권강민
- 2025년 2월 광주과학기술원 전기전자컴퓨터공학 학사
- 2025년 3월 ~ 현재 포항공과대학교 인공지능대학원 석사과정
- https://orcid.org/0009-0005-0165-4086

양희민
- 2022년 2월 포항공과대학교 컴퓨터공학과 학사
- 2022년 3월 ~ 포항공과대학교 컴퓨터공학과 통합과정
- https://orcid.org/0009-0004-2891-6093

조성현
- 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