저조도 비디오에서 Pseudo Ground Truth 생성을 위한 VBM3D와 SASSID 기반 파이프라인*

이경연[1], 이상연[1], 김준성[1], 이승연[2], 김성현[2], 정영민[2]
[1]서강대학교 컴퓨터공학과
[2]서강대학교 융합교육원, AI·SW교육센터
{joinme2000, lsy0163, holyrod, tmddus4671, kiles1201, ymchung}@sogang.ac.kr

A VBM3D-SASSID Pipeline for Pseudo Ground Truth Generation in Low-Light Video

Kyungyeon Lee[1], Sangyeon Lee[1], Junseong Kim[1], Seungyeon Lee[2], Sunghyeon Kim[2], YoungMin Chung[2]
[1]Department of Computer Science and Engineering, Sogang University, Seoul, Republic of Korea
[2]AI·SW Education Center, Sogang Institute for Convergence Education

요 약

저조도 비디오 영상에서는 낮은 신호 대 잡음비(SNR)로 인해 랜덤 노이즈와 banding noise와 같은 구조적 노이즈가 동시에 발생하여 영상 품질이 저하된다. 기존 영상 디노이징 방법은 전통적인 방식과 딥러닝 기반 방식으로 구분되며, 특히 딥러닝 기반 방법은 지도학습에 의존하여 깨끗한 ground truth(GT)를 필요로 한다. 그러나 저조도 환경에서는 GT 확보가 어려워 학습 기반 접근에 한계가 존재한다. 본 논문에서는 이러한 문제를 해결하기 위해, 저조도 비디오에서 pseudo ground truth를 생성하는 파이프라인을 제안한다. 제안 방법은 먼저 VBM3D를 적용하여 비디오의 시간적 정보를 활용해 노이즈를 완화하고, 이후 SASSID를 통해 self-supervised 방식으로 pseudo GT를 생성한다. 특히 SASSID 단독 적용 시 제거되지 않는 banding noise를 VBM3D를 통해 사전에 완화함으로써, 보다 안정적인 GT 생성이 가능하도록 한다.실험 결과, 제안한 방법은 기존 단일 방식 대비 향상된 시각적 품질과 정량적 성능을 보였으며, 저조도 비디오 디노이징을 위한 학습 데이터 생성에 효과적으로 활용될 수 있음을 확인하였다.

1. 서 론

저조도 환경에서 획득된 비디오 영상은 광자 수 감소로 인해 신호 대 잡음비(SNR)가 크게 저하되며, 이로 인해 랜덤 노이즈와 구조적 노이즈가 동시에 발생한다. 특히 구조적 노이즈는 영상 전반에 걸쳐 반복적인 패턴을 형성하며, 단순한 평균 기반 처리로는 제거가 어렵다 [1]. 영상 디노이징 방법은 크게 전통적인 방법과 딥러닝 기반 방법으로 구분된다. 전통적인 방법은 시간적 중복성과 영상의 자기 유사성을 활용하는 Rule-based 방식 [2, 3]으로 노이즈를 제거한다. 대표적인 전통 방식인 VBM3D [3]는 랜덤 노이즈에 효과적이지만 banding noise를 제외한 대부분의 구조적 노이즈를 완화하지 못한다.

딥러닝 기반 저조도 영상 복원 및 디노이징은 높은 복원 성능을 달성하였으나 [4], 대부분 지도학습에 의존하여 깨끗한 ground truth(GT)를 필요로 한다. 저조도 환경에서는 이러한 GT 확보가 어렵기 때문에, 이를 해결하기 위한 self-supervised 기반 방법 [5, 6]들이 제안되었다. 대표적으로 SASSID [6]는 spatially adaptive 특성을 활용하며, 랜덤 노이즈 및 일부 구조적 노이즈에 대해 효과를 보인다. 그러나 단일 프레임 기반으로 동작하여 비디오의 시간적 정보를 활용하지 못하며, 입력 데이터에 대한 의존성으로 인해 일반화 성능에 한계를 가진다. 또한 banding noise와 같은 전역적인 구조적 노이즈를 제거하지 못한다.

따라서 본 논문에서는 이러한 한계를 보완하기 위해, VBM3D와 SASSID를 결합한 파이프라인을 제안한다. 제안 방법은 먼저 VBM3D를 통해 다중 프레임 정보를 활용하여 전역적인 구조적 노이즈를 완화한 후, SASSID를 적용하여 pseudo ground truth를 생성한다. 이를 통해 비디오 기반 방법의 시간적 정보 활용 능력과 self-supervised 방법의 유연성을 결합하여, 저조도 비디오 환경에서 보다 안정적인 학습 데이터를 생성할 수 있다.

2. 관련 연구 및 연구 동기

영상 디노이징 연구는 크게 전통적인 방법과 딥러닝 기반 방법으로 구분되며, 특히 비디오 디노이징에서는 시간특성이 추가로 고려된다. 전통적인 방법은 이러한 시간 정보를 활용하여 노이즈를 제거하며, 대표적으로 VBM3D와 같은 다중 프레임 기반 알고리즘이 있다. VBM3D는 다양한 프레임에서 유사 패치를 탐색하고 이를 그룹화한 뒤, 변환 도메인 필터링과 aggregation을 통해 노이즈를 감소시키는 방식으로 동작한다. 이 과정에서 유사 패치 탐색은 프레임 간 움직임을 보정하는 효과를 가진다. 그러나 이러한 접근은 주로 랜덤 노이즈에 대한 제거 효과를 가지기 때문에 전역적인 구조적 노이즈인 banding 노이즈를 제외한 대부분의 구조적 노이즈는 완화가 어렵다.

딥러닝 기반 디노이징 방법은 데이터 기반 학습을 통해 높은 복원 성능을 달성하였으나, 대부분 지도학습에 의존하여 깨끗한 ground truth(GT)를 필요로 한다. 저조도 비디오 환경에서는 이러한 GT를 확보하기 어렵기 때문에, 이를 해결하기 위한 자기지도학습(self-supervised) 기반 방법들이 제안되었다. 대표적으

* 본 연구는 2026년 과학기술정보통신부 및 정보통신기획평가원의 AI중심대학 사업 지원을 받아 수행되었음(2026-0-00036)

로 SASSID는 spatially adaptive 특성을 활용하여, 랜덤 노이즈 및 일부 구조적 노이즈에 대한 효과적인 성능을 보인다. 그러나 SASSID는 단일 프레임 기반으로 동작하기 때문에 비디오의 시간적 정보를 활용하지 못하며, 입력 데이터에 대한 의존성으로 인해 일반화 성능에 한계를 가진다. 또한 banding noise와 같은 전역적인 구조적 노이즈는 제거하지 못하는 한계가 있다.

따라서 본 논문에서는 비디오 기반 전통적 방법과 self-supervised 방법의 한계를 보완하기 위해, VBM3D와 SASSID를 결합한 파이프라인을 제안한다. 제안 방법은 먼저 VBM3D를 통해 비디오의 시간적 정보를 활용하여 구조적 노이즈를 완화하고, 이후 SASSID를 적용하여 pseudo GT를 생성한다. 이를 통해 비디오 특화 알고리즘의 장점과 self-supervised 방법의 장점을 결합하여, 저조도 환경에서 보다 안정적인 학습 데이터를 생성할 수 있다.

3. 제안 방법

3.1 전체 파이프라인

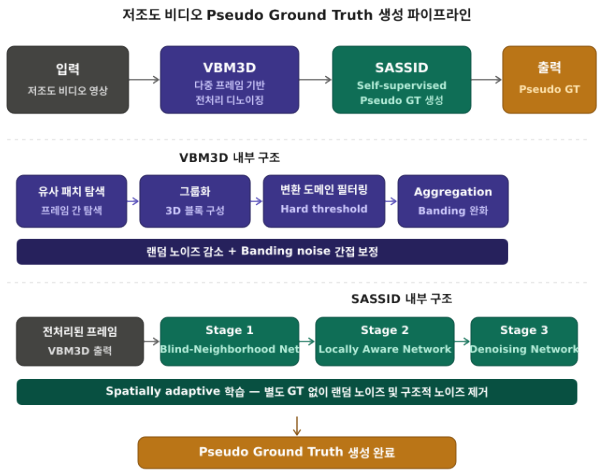

그림 1: 전체 파이프라인

본 논문에서는 저조도 비디오 영상으로부터 신뢰성 있는 pseudo ground truth(GT)를 생성하기 위한 두 단계 파이프라인을 제안한다. 저조도 환경에서는 광자 수 감소로 인해 랜덤 노이즈와 구조적 노이즈가 동시에 발생하며, 이를 단일 방법으로 제거하기 어렵다는 문제가 존재한다. 이에 본 연구는 다중 프레임 기반의 전통적 방법과 self-supervised 기반 학습 방법을 상호 보완적으로 결합하여 각 방법의 한계를 극복한다.

전체 구조는 그림 1와 같이 구성된다. 입력으로 저조도 비디오 영상을 받아, 1단계에서 VBM3D를 통해 다중 프레임 정보를 활용한 랜덤 노이즈 및 일부 구조적 노이즈 완화를 수행하고, 2단계에서 SASSID를 통해 self-supervised 방식으로 pseudo GT를 생성한다. VBM3D는 전역적인 구조적 노이즈인 banding noise를 사전에 완화함으로써 SASSID의 입력 품질을 향상시키고, SASSID는 별도의 clean GT 없이도 정밀한 노이즈 제거를 수행한다. 이러한 순차적 구조를 통해 단일 방법 대비 보다 안정적이고 고품질의 학습 데이터를 생성할 수 있다.

3.2 VBM3D 기반 전처리

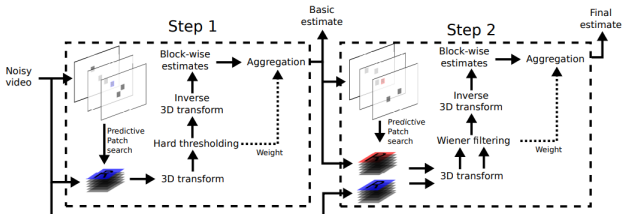

그림 2: VBM3D 알고리즘 아키텍처

VBM3D [3]는 비디오의 시간적 중복성과 영상의 자기 유사성(self-similarity)을 활용하여 노이즈를 제거하는 다중 프레임 기반 알고리즘이다. 그림 2에 제시된 바와 같이, VBM3D는 두 단계의 필터링 과정으로 구성된다. 1단계(Basic Estimate)에서는 Hard thresholding 기반의 변환 도메인 필터링을 수행하고, 2단계(Final Estimate)에서는 Wiener 필터링을 통해 보다 정밀한 복원을 수행한다.

본 연구에서는 VBM3D를 구조적 노이즈 완화를 위한 전처리 단계로 활용한다. VBM3D는 다양한 위치 및 프레임에서 유사한 패치를 탐색하고 이를 그룹화한 뒤, 3D 변환 도메인 필터링과 aggregation을 통해 노이즈를 감소시킨다. 유사 패치 탐색 과정은 프레임 간 작은 움직임을 간접적으로 보정하는 효과를 가지며, 안정적인 다중 프레임 집계를 가능하게 한다.

특히 aggregation 과정에서는 서로 다른 위치 및 프레임에서 수집된 패치들이 가중 평균되는데, 이 과정에서 row/column 방향으로 반복되는 banding noise의 bias 성분이 통계적으로 상쇄되는 효과가 발생한다. 단일 프레임 기반 방법은 이러한 구조적 패턴을 신호와 구분하지 못하지만, VBM3D의 다중 프레임 aggregation은 시간 축 정보를 활용하여 이를 간접적으로 완화할 수 있다. 이를 통해 이후 SASSID 단계에서 보다 안정적인 디노이징이 가능해진다.

3.3 SASSID 기반 Pseudo GT 생성

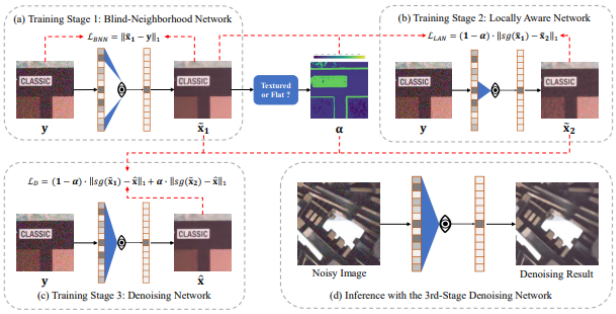

그림 3: SASSID 아키텍처

SASSID [6]는 self-supervised 기반의 단일 이미지 디노이징 방법으로, 영상의 평탄(flat) 영역과 질감(textured) 영역의 특성 차이를 고려하여 실사 sRGB 영상의 공간 상관 노이즈를 제거한다. 핵심 아이디어는 Stage 1에서 노이즈 상관 픽셀을 배제하

여 평탄 영역의 노이즈를 우선 제거하고, Stage 2에서 인접 픽셀을 활용해 질감 영역의 세부 디테일을 복원하는 것이다. 그림 3와 같이 세 단계로 구성되는데, Stage 1의 Blind-Neighborhood Network(BNN)는 blind 영역을 9×9로 확장하여 평탄 영역의 노이즈를 제거한 출력을 생성하고, Stage 2의 Locally Aware Network(LAN)는 인접 픽셀만을 활용하되 BNN 출력의 평탄 영역을 정답으로 삼아 학습함으로써 질감 영역의 디노이징 출력을 생성한다. Stage 3의 Denoising Network(U-Net)는 BNN과 LAN의 출력을 적응적 계수 α로 가중 합산한 값을 정답으로 삼아 학습되며, 추론 시에는 이 네트워크만 단독으로 사용된다.

SASSID는 입력 영상 내 정보만을 활용하여 학습을 수행하므로, 별도의 clean GT 없이도 노이즈 제거가 가능하다는 장점이 있다. 또한 영역별로 다른 복원 전략을 적용함으로써 랜덤 노이즈 및 일부 구조적 노이즈에 대해 효과적인 성능을 보인다. 그러나 SASSID를 단독으로 적용할 경우, banding noise와 같이 영상 전체에 걸쳐 일관된 방향성을 가지는 전역적 구조적 노이즈는 단일 프레임 내 정보만으로는 신호와 구분하기 어려워 충분히 제거되지 않는 한계가 존재한다.

본 연구에서는 이러한 문제를 해결하기 위해 VBM3D를 선행 단계로 적용한다. VBM3D를 통해 랜덤 노이즈와 banding noise가 사전 완화된 프레임을 SASSID의 입력으로 사용함으로써, SASSID가 추가적인 노이즈 제거에 집중할 수 있는 환경을 조성한다. 이와 같이 두 방법의 상호 보완적 결합을 통해, 저조도 비디오 환경에서도 시각적 품질과 정량적 성능이 모두 향상된 pseudo GT를 생성할 수 있다.

4. 실험 및 결과

4.1 실험 환경 및 데이터

본 연구에서는 저조도 비디오 디노이징 성능 평가를 위해 CRVD [7] 데이터셋을 사용하였다. CRVD는 실제 카메라 센서로 촬영된 RAW 비디오 기반 저조도 노이즈 데이터셋으로, 다양한 ISO 조건에서 획득된 noisy-clean 쌍을 포함한다. 특히 단일 이미지가 아닌 연속 프레임으로 구성되어 있어 시간적 정보를 활용하는 비디오 디노이징 방법의 성능을 평가하기에 적합하다.

본 실험에서는 입력 영상에 대해 VBM3D를 적용한 후, SASSID를 통해 pseudo ground truth를 생성하였다. 비교 대상으로는 VBM3D 단독 적용과 SASSID 단독 적용을 사용하였다.

4.2 정량적 평가

본 연구에서는 CRVD에서 제공되는 clean 영상을 기준으로 PSNR 및 SSIM을 계산하였다. 정량적 성능 평가를 위해 PSNR(Peak Signal-to-Noise Ratio)과 SSIM(Structural Similarity Index)을 사용하였다. 표 1는 CRVD 데이터셋에서의 평균 성능을 나타낸다. 제안한 방법은 VBM3D 및 SASSID 단독 적용 대비 PSNR과 SSIM 모두에서 향상된 성능을 보였다. 특히 두 방법의 장점을 결합함으로써 보다 안정적인 성능 향상이 나타남을 확인할 수 있다.

표 1: CRVD 데이터셋 평균 성능 비교

Method	PSNR (dB)	SSIM
VBM3D	35.8	0.923
SASSID	33.9	0.902
VBM3D + SASSID (Ours)	**36.9**	**0.933**

4.3 정성적 평가

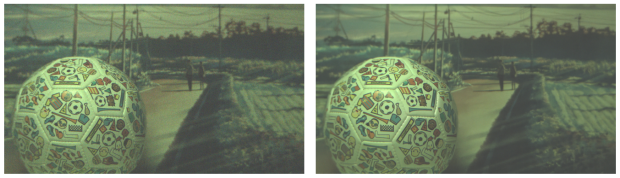

그림 4: 결과 이미지(CRVD scene7, 25600iso)

정성적 비교를 위해 입력 저조도 영상과 제안한 방법의 복원 결과를 비교하였다. 입력 영상에서는 전반적으로 미세한 노이즈가 존재하며, 특히 배경 영역에서 grain 형태의 노이즈가 관찰된다.

제안한 방법을 적용한 결과, 이러한 노이즈가 일부 완화되면서 보다 정돈된 시각적 결과를 확인할 수 있었다. 특히 배경 영역에서의 노이즈가 감소하여 전체적으로 보다 안정적인 영상 품질을 나타내며, 전경의 주요 객체 구조는 유지되는 경향을 보인다.

전반적으로 제안한 방법은 입력 영상 대비 노이즈를 완화하면서도 시각적 구조를 유지하는 방향으로 복원이 이루어짐을 확인할 수 있다.

참고 문헌

[1] K. Wei, Y. Fu, Y. Zheng, and J. Yang, "Physics-based noise modeling for extreme low-light photography," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 11, pp. 8520–8537, 2021.

[2] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on image processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[3] T. Ehret and P. Arias, "Implementation of the vbm3d video denoising method and some variants," *arXiv preprint arXiv:2001.01802*, 2020.

[4] C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to see in the dark," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3291–3300, 2018.

[5] J. Batson and L. Royer, "Noise2self: Blind denoising by self-supervision," in *International conference on machine learning*, pp. 524–533, PMLR, 2019.

[6] J. Li, Z. Zhang, X. Liu, C. Feng, X. Wang, L. Lei, and W. Zuo, "Spatially adaptive self-supervised learning for real-world image denoising," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9914–9924, 2023.

[7] H. Yue, C. Cao, L. Liao, R. Chu, and J. Yang, "Supervised raw video denoising with a benchmark dataset on dynamic scenes," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2301–2310, 2020.