

카메라 캘리브레이션 정보를 활용한 GAN 기반 차량용 서라운드뷰 생성 기법

강신재, 한동석*

경북대학교

kangsj129@knu.ac.kr, *dshan@knu.ac.kr

A GAN-Based Surround-View Generation Model for Vehicles Using Camera Calibration Information

Shin Jae Kang, Dong Seog Han*

School of Electronic and Electrical Engineering, Kyungpook National Univ.

요약

본 논문은 다중 어안 카메라 입력으로부터 BEV(bird's-eye view) 서라운드뷰 영상을 생성하기 위해, 카메라 캘리브레이션 정보를 학습에 반영한 GAN(generative adversarial network) 기반 BEV 생성 기법을 제안한다. 제안 방법은 카메라 내·외부 파라미터에서 유도된 기하 정보들을 특정 융합 단계에 반영하여, 다중 카메라 특징을 BEV 좌표계 기준으로 정렬하고 시점 차이로 인한 위치 불일치와 경계부 불연속을 완화한다. 또한 재구성 손실, 에지 보존 손실, 지각 손실, 특정 일치 손실을 결합한 학습 전략을 통해 BEV 전반의 구조적 정합과 경계부 품질을 동시에 개선한다. FB-SSEM 합성 데이터셋에서 실험한 결과, 제안 방법은 PSNR(peak signal-to-noise ratio) 30.58 dB를 기록하여 기존 BevGAN 대비 약 7.5 dB, BridgeGAN 대비 약 4.3 dB 이상의 성능 향상을 보였다. 또한 L1 오차는 0.02로 감소하였고, SSIM(structural similarity index measure)은 0.89로 구조적 유사성이 크게 개선되었다. 더불어 SD(seam discontinuity)와 LPIPS(learned perceptual image patch similarity)의 유의미한 감소는 카메라 접합부에서 발생하는 경계부 불연속과 지각적 불일치가 효과적으로 억제되었음을 보여준다.

I. 서론

차량 주변 환경을 직관적으로 표현하는 BEV(bird's-eye view)는 주차 및 저속 주행 보조 시스템에서 핵심적인 시각화 수단으로 활용되고 있다. 현재 상용 BEV 시스템은 다수의 어안 카메라 영상을 IPM(inverse perspective mapping)으로 투영하여 생성하지만, 장면을 단일 평면으로 가정하는 기하학적 제약으로 인해 높이를 가진 객체에 대해 심각한 형태 왜곡이 발생한다 [1]. 또한 다중 카메라 영상이 결합되는 과정에서 시점 차이(parallax)와 가림 현상(occlusion)에 의해, 경계부에서 객체의 위치 불일치, 고스팅(ghosting)과 같은 시각적 불연속이 빈번히 나타난다. 이러한 현상은 객체의 실제 형태와 공간적 연속성에 대한 인지를 저해하며, 사람이 직접 확인하는 BEV 영상의 신뢰성과 시각적 일관성을 크게 제한한다.

이러한 기하 기반 방식의 한계를 완화하기 위해 최근에는 학습 기반 BEV 생성 기법이 제안되고 있다 [2][3]. 이들 방법은 다중 카메라 간 중첩 영역에서의 특징 융합과 경계부 처리를 데이터 기반으로 학습함으로써, 평면 투영에 의존하지 않는 보다 자연스러운 BEV 시각화를 목표로 한다. 본 논문은 이러한 접근을 바탕으로, 다중 어안 카메라 입력과 카메라 캘리브레이션 정보를 학습 과정에 반영하여 다중 카메라 간 기하 관계를 보다 일관되게 활용하고, 다중 판별기와 다양한 손실 함수를 통해 왜곡 및 경계부 시각적 불연속을 효과적으로 완화한 고품질 BEV 영상 생성을 목표로 한다.

II. 캘리브레이션 정보를 반영한 GAN 기반 BEV 생성 네트워크

추정된 카메라 내부 파라미터와 차량 좌표계 기준 외부 파라미터를 이용

하여, 그림 1에서 제안하는 BEV 생성 네트워크의 구조는 다중 어안 카메라 입력으로부터 BEV 서라운드뷰 영상을 생성한다. 각 카메라 뷰는 공유된 CNN(convolutional neural network) 백본을 통해 다양한 해상도의 특징으로 인코딩되며, 이후 트랜스포머 기반 융합 모듈에서 교차 어텐션(cross-attention)을 사용하여, 각 카메라에서 추출된 다중 해상도 특징을 BEV 좌표계 기준으로 정렬하고 하나의 통합된 특징 표현으로 결합한다.

이 과정에서 카메라 내부 및 외부 파라미터로부터 유도된 캘리브레이션 정보들을 기반으로, 어안 카메라의 기하학적 특성을 반영한 위치 임베딩(embedding)을 구성한다. 해당 임베딩은 각 픽셀의 광선 방향과 차량 좌표계와의 관계를 표현하며, 어안 렌즈의 비선형 왜곡 특성을 포함한다. 이러한 위치 임베딩은 특정 융합 단계에서 각 카메라 뷰의 특징과 결합되어 사용된다. 이 임베딩은 카메라 간 시점 차이로 인한 위치 불일치와 시차를 완화하여, 다중 카메라 특징을 BEV 좌표계 기준으로 정렬하고 중첩 영역에서도 안정적인 특징 결합을 가능하게 한다.

정렬된 BEV 특징은 경량화된 복원 헤드를 통해 최종 RGB 서라운드뷰 영상으로 변환된다. 그림 2에 나타난 바와 같이 BEV 공간에서의 구조를 안정적으로 복원하고 경계부 품질을 향상시키기 위해, 서로 다른 역할을 수행하는 여러 손실 항을 함께 사용한다. 재구성 손실은 BEV 전반의 구조적 정합을 유지하도록 유도하며, 에지 보존 항은 차선이나 객체 윤곽과 같은 경계 정보를 선명하게 유지하는 역할을 한다. 지각 손실은 생성된 영상이 시각적으로 자연스러운 분포를 따르도록 보조하며, 특정 일치 항은 학습 과정의 불안정을 완화하여 안정적인 수렴을 돋는다. 또한 다중 판별기를 통한 적대적 손실은 BEV 영상의 전역적 구조와 경계부 국소 영역을 동시에 제어하여, 경계부 불연속이 억제된 일관된 서라운드뷰 생성을 가능하게 한다.

실험의 경우 FB-SSEM 합성 데이터셋에서 전·후·좌·우 네 방향의 어안 카메라 입력과 이에 대응하는 BEV RGB 이미지를 감독 신호로 사용하였다. 학습은 400 에포크, 배치사이즈 1 설정에서 진행되었고, Adam 옵티마이저를 사용하였다. 생성기와 판별기의 학습률은 각각 0.001과 0.0001로 설정하였으며, 모든 비교 실험은 동일한 학습 조건에서 수행되었다.

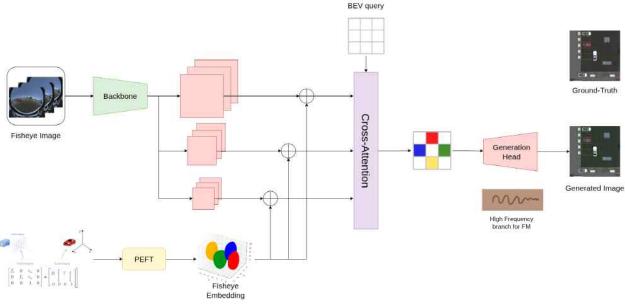


그림 1. 딥러닝 기반 BEV 생성 네트워크

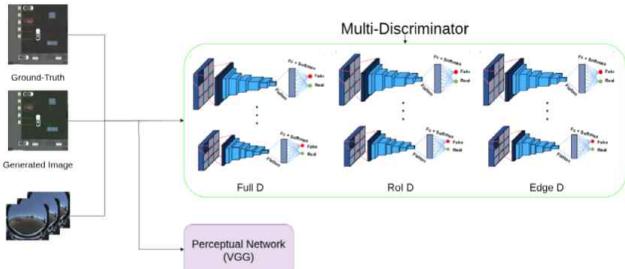


그림 2. BEV 생성 네트워크의 다중 판별기 구조

표 1에서 보듯이, 제안 방법은 모든 정량 지표에서 기존 방법들을 큰 폭으로 상회한다. PSNR(peak signal-to-noise ratio)은 30.58로 BevGAN 대비 약 7.5 dB, BridgeGAN 대비 약 4.3 dB 이상 향상되었으며, L1 오차 또한 0.02로 크게 감소하였다 [2][3]. SSIM(structural similarity index measure)은 0.89로, 기존 방법 대비 구조적 유사성이 현저히 개선되었음을 보여준다. 또한 SD(seam discontinuity)와 LPIPS(learned perceptual image patch similarity)의 큰 감소는 경계부 불연속과 텍스처(texture) 왜곡이 효과적으로 억제되었음을 의미한다. 이는 제안 방법이 전역 구조와 경계부 품질을 동시에 개선함을 정량적으로 입증한다.

그림 3은 BEV 생성 결과에 대한 정성적 비교를 보여준다. (a)는 GT이며, (b)는 BevGAN, (c)는 BridgeGAN_homo, (d)는 BridgeGAN, (e)는 제안 방법의 결과이다. 기존 방법들인 BevGAN과 BridgeGAN 계열은 전반적인 구조는 유지하지만 노면 텍스처, 차선, 객체 윤곽이 상대적으로 희미하게 표현된다. 반면 제안 방법은 고주파 성분과 세부 디테일을 충실히 반영하여, 시각적으로 GT(ground truth)에 가장 근접한 BEV 이미지를 생성함을 확인할 수 있다.

표 1. 서라운드뷰 생성 모델 정량적 비교

Method	PSNR	L1	SSIM	SD	LPIPS
BevGAN[2]	23.05	0.14	0.42	0.23	0.73
BridgeGAN _{homo} [3]	25.67	0.10	0.35	0.27	0.73
BridgeGAN[3]	26.24	0.095	0.36	0.27	0.70
Ours	30.58	0.02	0.89	0.05	0.10

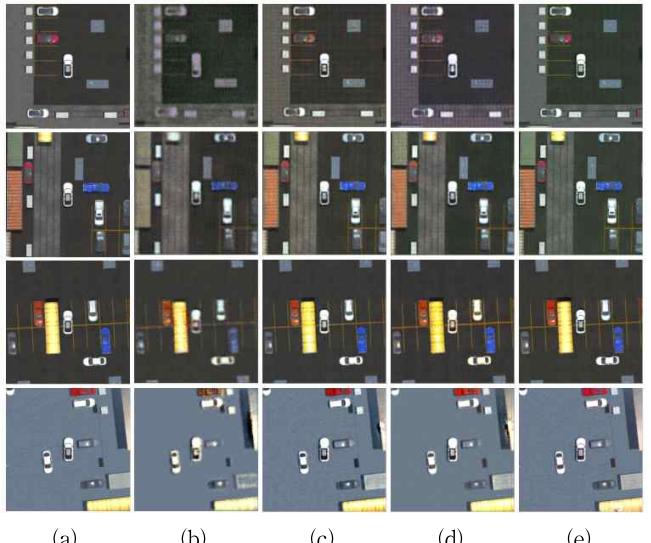


그림 3. BEV 생성 이미지의 정성적 비교

III. 결론

본 논문에서는 IPM 기반 BEV 생성 방식이 갖는 평면 가정의 한계로 인해 발생하는 높이 객체 왜곡, 시점 차이 및 가림 현상으로 인한 시각적 불연속 문제를 지적하고, 이를 완화하기 위한 학습 기반 BEV 생성 방법을 제안하였다. 또한 카메라 내부·외부 캘리브레이션 정보를 학습 과정에 반영함으로써 기하 구조와 시각적 품질을 동시에 고려한 BEV 생성을 가능하게 한다.

정성적·정량적 실험 결과, 제안 방법은 기존 방법 대비 노면 텍스처와 차선, 객체 윤곽과 같은 세부 디테일을 보다 선명하게 복원하며, 시각적으로 정답 BEV 이미지에 가장 근접한 BEV 영상을 생성함을 확인하였다. 이는 학습 기반 시각화에 기하 정보를 결합함으로써, FB-SSEM 합성 데이터셋에서 신뢰도 높은 서라운드뷰 BEV 시각화를 구현할 수 있음을 보여준다. 향후 연구에서는 추출된 카메라 캘리브레이션 정보를 통해 실제 주행 환경에서의 적용 가능성으로 확장할 계획이다.

ACKNOWLEDGMENT

이 논문은 2021년도 교육부 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (RS-2021-NR060127)

참 고 문 헌

- J. Phlion and S. Fidler, "Lift, Splat, Shoot: Encoding Images From Arbitrary Camera Rigs by Implicitly Unprojecting to 3D," in Proc. European Conference on Computer Vision (ECCV), 2020, pp. 194 - 210.
- R. Benaissa, A. Musabini, R. Benmokhtar, M. Bakthavatchalam, and X. Perrotton, "BevGAN: Generative fisheye cross-view transformers," in Proceedings of the 17th International Conference on Agents and Artificial Intelligence - Volume 2: ICAART, INSTICC. SciTePress, 2025, pp. 153 - 163.
- X. Zhu, Z. Yin, J. Shi, H. Li, and D. Lin, "Generative adversarial frontal view to bird view synthesis," in 2018 International Conference on 3D Vision (3DV), 2018, pp. 454 - 463..