

저조도 주행 환경에서 확산 모델 기반 이미지 품질 개선을 통한 카메라 궤적 추정 성능 향상 연구

김승욱¹, 김준오², 조경은³

¹동국대학교 컴퓨터·AI학과 석사과정

²동국대학교 NUI/NUX 플랫폼 연구센터 교수

³동국대학교 컴퓨터·AI학부 교수

swk1067@naver.com, junoh.kim@dongguk.edu, cke@dongguk.edu(교신저자)

A Diffusion Model-Based Image Enhancement for Improved Camera Trajectory Estimation in Low-Light Driving Environments

Seungwook Kim¹, Junoh Kim², Kyungeun Cho³

¹Dept. of Computer Science and Artificial Intelligence, Dongguk University

²NUI/NUX Platform Research Center, Dongguk University

³Dept. of Computer Science and Artificial Intelligence, Dongguk University

요 약

단안 카메라 기반 차량 궤적 추정에서 조도, 날씨 변화 등의 요인으로 인한 이미지 품질 저하는 Visual Odometry 성능에 영향을 미친다. 본 연구는 특징점 기반 및 딥러닝 기반 카메라 궤적 예측 모델의 성능 향상을 위해 이미지 품질 개선 방법을 접목한 파이프라인을 제안한다. 확산 모델 기반 이미지 해상도 향상 기법을 전처리 모듈로 활용하여, 기존 모델에 모듈식으로 통합할 수 있다. nuScenes 저조도 데이터를 활용한 실험에서 특징점 기반 ORB Visual Odometry에서 평균 42.9% ATE 개선을 보였다. MonoDepth2에서 학습된 PoseNet에서는 평균 4.6%의 개선을 확인하였다.

1. 서론

자율주행 연구 분야에서 차량의 정확한 이동 경로 추정은 핵심 기술 중 하나이다. 단안 RGB 카메라는 비용 효율성과 높은 활용성 때문에 널리 사용되지만, 3차원 정보가 제공되지 않아 거리(distance) 추정에 본질적인 한계를 지닌다.[1] 이러한 환경에서 연속된 이미지 프레임을 이용해 카메라의 움직임을 추정하는 Visual Odometry(VO) 기술은 카메라 성능과 인공지능으로 인해 발전을 거듭하고 있다.

대표적인 VO 방식으로는 ORB 특징점 기반의 기하학적 방법[2]과 딥러닝 기반의 Feature matching 연구[3]가 있다. 하지만 두 방식 모두 명확한 거리가 없을 경우, 입력 이미지의 특징에 의존하여 Shift & Affine-transform의 문제로 정의한다. 특히 저조도, 악천후와 같은 특징 추출의 일관성 유지가 어려운 환경에서는 이미지 품질 저하로 인해 특징점 추출 및 매칭이 불안정해지고, 이는 카메라 궤적 추정 정확도를 떨어뜨리는 주요 원인이 된다.

본 연구는 이러한 배경에서, 기존 VO 모델의 구조 변경이나 재학습 없이 적용이 가능한 파이프라인

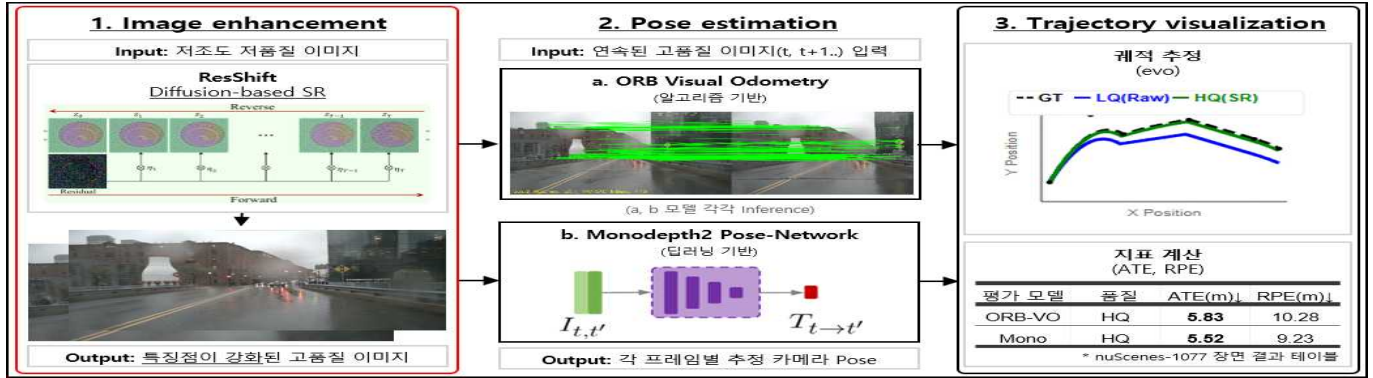
을 제안한다. 확산 모델을 기반으로 빠르게 이미지 품질을 높이는 ResShift[4]를 기반으로 전처리 모듈을 적용하여 저조도 주행 환경에서의 이미지 품질을 개선함으로써, 기존 VO 모델의 카메라 궤적 추정 성능을 높이는 연구이다.

2. 관련 연구

2.1 이미지 품질 개선

이미지 품질 개선 기술은 품질이 저하된 이미지로부터 원본의 시각 정보를 복원하는 것을 목표로 한다. 이는 크게 수학적 기법과 인공지능 기반 기법으로 나눌 수 있다. 먼저, 해상도 향상을 위한 양선형(Bilinear)이나 양삼차(Bicubic) 보간법과 같은 전통적 방식은 속도가 빠르지만, 세부 정보 복원이 어려워 이미지가 흐릿해지는 한계가 있다. 반면, 인공지능 기반 합성곱 신경망(CNN)을 적용한 초고해상도(Super-Resolution, SR) 기법들은 더 선명하고 사실적인 결과를 생성하며 주류 기술로 자리 잡았다.[5]

최근에는 이미지 생성 분야에서 우수한 성능을 보인 확산 모델(Diffusion Model)이 이미지 복원 연구



(그림 1) 전체 프레임워크

에서도 주목받고 있다. 확산 모델은 노이즈로부터 이미지를 점진적으로 복원하는 과정을 통해 높은 품질의 결과를 생성하지만, 수많은 반복 샘플링 과정으로 인해 추론 속도가 느리다는 단점이 있다.[6]

한편, 저조도 주행 영상은 낮은 품질로 특징점 추출이 어려워 Visual Odometry 모델 추정 정확도가 저하된다. ResShift[4]는 기존 Diffusion 모델의 속도 한계를 극복한 기법으로, 고품질 이미지와의 잔차(residual)를 점진적으로 이동시키는 마르코프 체인을 설계하여 단 4단계의 짧은 샘플링만으로도 고품질 복원을 달성한다. 합성 노이즈뿐 아니라 실제 저조도 환경 촬영 이미지에서도 강건한 초고해상도(SR) 성능을 보인다. 따라서 ResShift에 추가적인 학습을 하지 않고도, 기존 VO 모델에 바로 접목이 가능하다는 실용적 장점이 있다.

2.2 Visual Odometry

ORB Visual Odometry[2]는 ORB(Oriented FAST and Rotated BRIEF) 특징점을 활용하여 연속된 이미지 프레임 간의 대응점을 추출 및 매칭하고, RANSAC(Random Sample Consensus) 알고리즘으로 기하학적으로 유효한 대응점을 선별하여 카메라의 움직임을 추정하는 방식이다. Essential Matrix 계산을 통해 카메라의 회전과 이동을 추정하여 전체 궤적을 계산한다. ORB 방식은 빠른 처리 속도와 구현의 용이성으로 널리 사용되지만, 이미지 특징에 의존하여 품질에 민감하다는 특성이 있다.

MonoDepth2[3]는 깊이와 카메라 자세(pose)를 추정하는 대표적인 자기지도(self-supervised) 딥러닝 모델이다. 깊이와 자세 추정은 ResNet18 기반으로 독립적인 네트워크에서 이루어지며, pose 추정 네트워크는 연속 프레임으로부터 카메라 자세와 움직임을 추정한다. Monodepth는 KITTI 데이터셋[7]으로 사전 훈련된 모델을 제공하여 nuScenes[8]과 같은

다른 도심 주행 데이터셋에서도 제로샷(zero-shot) 성능을 평가할 수 있다.

3. 제안하는 기법

3.1 전체 프레임워크

제안하는 기법의 프레임워크는 (그림 1)과 같다. 이미지 품질 개선, Pose 추정을 위한 Pair images 생성 및 전달, 카메라 궤적 시각화 및 평가 모듈을 순차적으로 연결한 구조를 가진다.

3.2 이미지 품질 개선

이미지 품질 개선은 ResShift 모델을 이용한 초고해상도 단일 단계로 구성된다. 저조도 환경에서 상대적으로 부족한 경계선과 같은 세부 정보를 효과적으로 복원하고 노이즈를 감소시키는 일종의 정제(refinement) 효과를 가진다. 이를 통해 후속 단계인 특징점 검출에 유리한 고품질 이미지를 생성한다.



(그림 2) ResShift SR을 통한, 품질 개선 전/후 이미지

3.3 Pose 추정 및 궤적 시각화

생성된 고품질 이미지는 ORB-VO와 MonoDepth2 Pose Network의 입력 데이터로 각각 사용되어 연속된 프레임 간 카메라의 상대적인 자세(회전, 이동)를 추정한다. 추정된 상대 자세들을 순차적으로 누적하여 카메라가 장착된 차량의 이동 궤적을 복원한다.

마지막으로, evo 라이브러리를 활용하여 추정된 이동 궤적을 실제 이동 궤적(Ground Truth, GT)과 비교하고, ATE와 RPE 같은 정량적 평가 지표를 계산하여 나온 성능을 평가한다.

4. 실험

4.1 개요

OS (GPU): Linux Ubuntu 24.04 (RTX4070 Ti)

데이터셋: nuScenes[8] 데이터셋에서 실제 카메라 경로가 있는 도심 주행 장면 중, 야간(Scene-1077, 1003), 해질녘(Scene-0996), 우천(Scene-0450) 상황에서 촬영된 전방 카메라 이미지를 사용하였다.

평가 지표: 궤적 추정의 전역적 정확도를 평가하기 위해 절대 궤적 오차(ATE)를 사용하였다. ATE는 추정 궤적과 실제 궤적(GT) 간 차이를 RMSE(Root Mean Square Error)로 계산한 값으로, 값이 작을수록 추정 성능이 우수함을 나타낸다.

4.2 정량적 결과

야간 주행 장면(Scene-1077)에서 궤적 추정 비교 결과는 <표 1>과 같다. 약 1,000회의 RANSAC Iteration을 거친 ORB-VO의 경우, 원본을 사용했을 때의 ATE는 12.57m였으나, 제안하는 기법으로 개선된 이미지를 사용했을 때는 5.83m로 오차가 약 53.7% 감소하는 큰 성능 향상을 보였다. 딥러닝 기반의 MonoDepth2 모델 역시 ATE가 5.71m에서 5.52m로 약 3.3% 개선되어, 제안하는 기법의 효과를 확인할 수 있었다. KITTI로 사전 훈련된 mono+stereo_1024x320 가중치를 사용하였다.

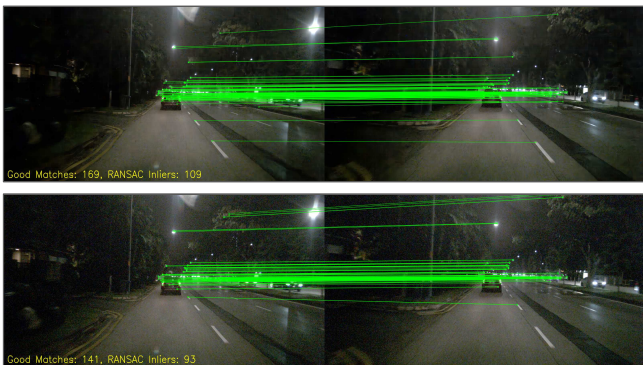
<표 1> Scene-1077 궤적 추정 결과 비교

평가 모델	이미지 품질	ATE(m) ↓	RPE(m) ↓
ORB-VO	LQ(Raw)	12.57	10.28
	HQ(SR)	5.83	10.28
MonoDepth2	LQ(Raw)	5.71	9.22
	HQ(SR)	5.52	9.23

*LQ: Low Quality, HQ: High Quality

<표 2> ORB-VO 특징점 매칭 평균값 비교

	LQ(Raw)	HQ(SR)	개선율 ↑
Good Matches ↑	334	363	+8.7%
RANSAC Inliers ↑	239	257	+7.5%



(그림 3) 1077 장면 특징점 매칭 비교

*RANSAC 대응점: 위(HQ) 109개, 아래(LQ) 93개

이러한 성능 향상의 직접적인 원인을 분석하기 위해 특징점 매칭 성능을 비교한 결과는 <표 2>이다. 이미지 품질 개선 후, Lowe's ratio test를 통과한 유효 특징점(Good Matches)의 수는 평균 8.7%, RANSAC을 통해 기하학적으로 일관성이 검증된 최종 대응점(RANSAC Inliers) 수는 평균 7.5% 증가하였다. 이는 더 많고 신뢰도 높은 특징점이 카메라 자세(Pose) 추정의 정확도를 높였음을 보여준다.

4.3 다중 장면 검증 결과

제안하는 기법의 일반성을 검증하기 위해 해질녘, 우천 등 다양한 환경에서 추가 실험을 진행하였다. 비교 결과는 <표 3, 4>와 같다. 주간은 큰 변화가 없었으나, 저조도 환경에서 일관되게 ATE가 감소하였으며, 우천 시 46.7%의 ATE 개선율을 보였다.

<표 3> 장면별 ORB-VO 궤적 추정 결과 비교

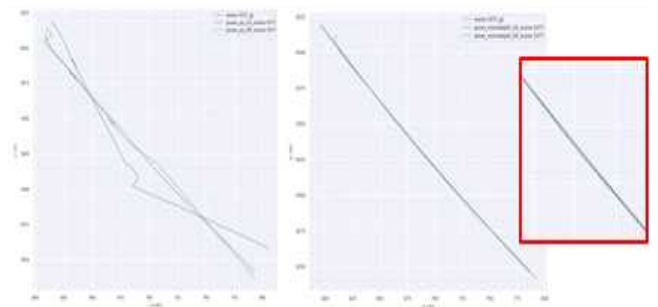
장면	환경	LQ		HQ		ATE 개선율 ↑
		ATE ↓	RPE ↓	ATE ↓	RPE ↓	
0916	주간	2.01	3.34	2.00	3.29	0.50%
0450	우천	2.40	5.10	1.28	5.04	46.7%
0996	해질녘	4.71	2.72	4.60	2.65	2.3%
1003	야간	9.66	5.65	3.03	5.83	68.6%

<표 4> 장면별 MonoDepth2 궤적 추정 결과 비교

장면	환경	LQ		HQ		ATE 개선율 ↑
		ATE ↓	RPE ↓	ATE ↓	RPE ↓	
0916	주간	6.77	8.89	6.69	8.82	1.2%
0450	우천	3.38	5.26	3.29	5.27	2.7%
0996	해질녘	3.13	2.68	3.04	2.69	2.9%
1003	야간	3.02	5.15	2.73	5.18	9.6%

4.4 정성적 결과

(그림 4)는 야간 환경(scene-1077)에서 두 모델로 추정된 궤적을 실제 궤적(GT, 점선)과 비교한 결과이다. ORB에서 원본(LQ, 파란색)은 궤적 오차가 있는 반면에, 고품질 이미지(SR, 녹색)의 추정 궤적은 실제 궤적(nuScenes GT)과 근접한 것 확인할 수 있다. MonoDepth는 큰 궤적의 변화는 없었으나, 확대하여 보면 기존보다 개선된 모습을 볼 수 있다.



(그림 4) 궤적 추정 결과(좌: ORB, 우: MonoDepth2)



(그림 5) scene-0450(우천), ORB VO 궤적 추정 결과

* 파랑선: LQ, 초록선: SR(개선된 pose), 검정선: GT

5. 결론

본 연구는 저조도 주행 환경에서 이미지 품질 개선을 통해 Visual Odometry의 성능을 향상시키는 방법을 제안하였다. ResShift 기반 전처리만으로 nuScenes 데이터의 저조도 주행 환경에서 특징점 기반 ORB-VO의 ATE를 평균 42.9%, MonoDepth2의 ATE를 4.6% 개선하였다. (산술 평균)

본 연구의 기여는 크게 3가지이다. (1) 기존 학습된 모델의 재학습 없이 야생(wild) 환경의 저품질 데이터를 전처리 단계에서 정제하여 VO 성능을 향상시키는 실용적 방법론을 제시하였다. (2) 특징점 기반과 딥러닝 기반 두 방법론 모두에서 이미지 품질 개선의 효과를 검증하여, 저비용 단안 카메라만으로도 고성능 VO가 가능함을 실증하였다. (3) 다양한 저조도 환경에서 일관된 성능 향상을 보여 제안 방법의 강건성과 범용성을 확인하였다.

향후 연구로는 실시간 처리가 가능하도록 경량화된 이미지 품질 개선 모델을 연구하고, 더욱 다양한 VO 및 SLAM 알고리즘과의 호환성을 검증하는 방향으로 확장할 수 있을 것이다.

사사

* 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. RS-2022-NR070225).(50%)

* 본 연구성과물은 2025년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. RS-2025-25432196).(45%)

* 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 인공지능융합혁신인재양성사업 연구 결과로 수행되었음(IITP-2025-RS-2023-00254592).(5%)

참고문헌

- [1] Scaramuzza, D., & Fraundorfer, F., Visual odometry [tutorial], IEEE robotics & automation magazine, 18, 4, 80-92, 2011.
- [2] Mur-Artal, R., Montiel, J. M. M., & Tardós, J. D., ORB-SLAM: a Versatile and Accurate Monocular SLAM System, IEEE Transactions on Robotics, 31, 5, 1147-1163, 2015.
- [3] Godard, C., Mac Aodha, O., Firman, M., & Brostow, G. J., Digging into Self-Supervised Monocular Depth Estimation, IEEE International Conference on Computer Vision, Seoul, 2019, 3828-3838.
- [4] Yue, Z., Wang, J., & Loy, C. C., ResShift: Efficient diffusion model for image super-resolution by residual shifting, Advances in Neural Information Processing Systems, New Orleans, 2023, 12812-12825.
- [5] Dong, C., Loy, C. C., He, K., & Tang, X., Learning a deep convolutional network for image super-resolution, European Conference on Computer Vision, Zurich, 2014, 184-199.
- [6] Ho, J., Jain, A., & Abbeel, P., Denoising diffusion probabilistic models, Advances in Neural Information Processing Systems, 2020, 6840-6851.
- [7] Geiger, A., Lenz, P., Stiller, C., & Urtasun, R., Vision meets robotics: The KITTI dataset, International Journal of Robotics Research, 32, 11, 1231-1237, 2013.
- [8] Caesar, H., Bankiti, V., Lang, A. H., Vora, S., Liong, V. E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., & Beijbom, O., nuScenes: A multimodal dataset for autonomous driving, IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, 2020, 11621-11631.