

영상 오도메트리 성능 향상을 위한 파티클 필터의 어텐션 기반 리샘플링 기법

Attention-based Particle Resampling in Sequential Importance Resampling Filters for Enhanced Visual Odometry Performance

강 창 호¹, 김 선 영^{2,*}

(Chang Ho Kang¹ and Sun Young Kim^{2,*})

¹Department of AI Robotics, Sejong University

²School of Mechanical Engineering / Advanced Machinery Technology Research Institute, Kunsan National University

Abstract: Accurate state estimation is essential for autonomous driving in global navigation satellite system (GNSS)-denied environments. Particle filters, particularly sequential importance resampling (SIR), are widely used for sensor fusion but suffer from weight degeneracy and particle impoverishment during resampling. To address this, we propose Particle Transformer, a deep learning-based resampling framework combining weighted multi-head attention for particle importance modeling and a variational autoencoder (VAE) for diversity preservation. An adaptive strategy based on effective sample size (ESS) further balances accuracy and efficiency. Experiments on the KITTI odometry dataset show 3.0 % lower absolute trajectory error (ATE) and 21.2 % lower relative pose error (RPE) compared to systematic resampling, while maintaining maximum diversity with only 6.4 ms per frame.

Keywords: autonomous driving, visual odometry, particle filter, attention, particle resampling, sequential importance resampling filter

I. 서론

자율주행 자동차가 도심 협곡, 터널, 실내 주차장 등 위성 항법 시스템(GNSS, Global Navigation Satellite System)의 신호 수신에 불가능하거나 신뢰도가 낮은 환경에서 안정적으로 주행하기 위해서는 차량의 현재 위치와 자세를 정밀하게 추정하는 기술이 필수적이다[1-5]. 이를 위해 카메라, 라이다(LiDAR), 관성 측정 장치(IMU, Inertial Measurement Unit) 등 다양한 센서 정보를 융합하는 연구가 활발히 진행되고 있으며 [6-8], 비선형(non-linear) 및 비가우시안(non-Gaussian) 시스템에서 뛰어난 성능을 보이는 파티클 필터(particle filter), 특히 순차적 중요도 리샘플링(SIR, Sequential Importance Resampling) 필터가 널리 사용되고 있다[9-11].

파티클 필터는 상태 공간의 확률 분포를 다수의 파티클 집합으로 근사하여 추정하는데[12], 필터가 반복됨에 따라 소수의 파티클에만 가중치가 집중되는 가중치 퇴화(weight degeneracy) 현상이 필연적으로 발생한다. 이를 해결하기 위해 가중치가 낮은 파티클을 제거하고 가중치가 높은 파티클을 복제하는 리샘플링(resampling) 과정이 수행된다. 그러나 기존의 리샘플링 방법들은 파티클의 가중치 크기만을 고려한 단순 복제 방식을 취하기 때문에, 파티클 집합의 다양성(diversity)이 급격히 감소하는 샘플 빈곤화(sample impoverishment) 문제를 야기한다. 이는 급격한 회전이나 가속과 같은 동적

상황에서 추정 성능을 저하시키고 국소 최적해로 조기 수렴(premature convergence)하는 원인이 된다.

이러한 한계를 극복하기 위해 다양한 리샘플링 기법들이 연구되어왔다. 가장 기본적인 다항(multinomial) 리샘플링은 가중치에 비례하여 파티클을 무작위로 복원 추출하는 방식으로 구현이 간단하지만, 샘플링 분산이 커서 파티클 다양성을 훼손할 위험이 크다. 이를 개선하기 위해 제안된 체계적(systematic) 리샘플링은 누적 가중치 분포에서 균등한 간격으로 파티클을 선택하여 분산을 최소화하고, $O(N)$ 의 빠른 연산 속도를 제공하여 현재 가장 널리 사용되고 있다. 분산을 줄이기 위한 또 다른 접근법으로 층화(stratified) 리샘플링과 잔차(residual) 리샘플링이 있다. 층화 리샘플링은 전체 구간을 파티클 수만큼 나누어 각 구간에서 파티클을 선택하며, 잔차 리샘플링은 가중치에 비례하여 확정적으로 파티클을 선택한 후 남은 부분을 무작위로 채우는 방식이다. 이들은 체계적 리샘플링보다는 유연하지만 여전히 높은 가중치를 가진 파티클을 단순 복제한다는 한계에서 벗어나지 못한다.

파티클의 다양성을 인위적으로 늘리기 위한 시도로 정규화(regularized) 리샘플링이 제안되었다. 이는 커널 밀도 추정(kernel density estimation)을 통해 연속적인 확률 분포를 생성한 후 리샘플링을 수행하여 파티클이 겹치지 않게 하려 했으나, 커널 대역폭 설정에 민감하고 계산 비용이 높다는 단점이 있

* Corresponding Author

Manuscript received December 4, 2025; revised January 10, 2026; accepted January 27, 2026

강창호: 세종대학교 AI로봇학과 조교수(kcguri@sejong.ac.kr, ORCID[®] 0000-0002-9899-3076)

김선영: 국립군산대학교 기계공학부 부교수(sykim77@kunsan.ac.kr, ORCID[®] 0000-0002-6189-0268)

※ 본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(No. RS-2025-24534687)과 2025년도 교육부 및 서울특별시의 재원으로 서울RISE센터의 지원을 받아 수행된 지역혁신중심 대학지원체계(RISE)의 결과임(2025-RISE-01-019-04).

※ 이 논문은 2025 ICROS-KROS 호남제주 그랜드 합동 학술대회에 초안이 발표되었음.

다. 나아가 최적 수송(optimal transport) 이론을 적용하여 파티클 분포 간의 거리를 최소화하는 방식도 연구되었으나 [13], 이 또한 높은 연산량을 요구하여 실시간성에 제약이 있다.

최근 딥러닝 기술의 발전으로 트랜스포머(Transformer)와 같은 어텐션 기반 모델이 다양한 분야에서 성과를 보이고 있다[14,15]. 본 논문에서는 이러한 딥러닝 기술을 파티클 필터의 리샘플링 단계에 접목하여 기존 방법론의 한계를 극복하고자 한다[16]. 제안하는 방법은 기존의 통계적 기법들과 달리, 데이터 기반의 학습을 통해 파티클 간의 상호작용과 중요도를 명시적으로 모델링한다. 본 연구의 주된 기여는 다음과 같다.

- 파티클 트랜스포머 제안: 가중치가 반영된 멀티헤드 어텐션(weighted multi-head attention) 메커니즘을 통해 파티클 간의 상호작용과 중요도를 명시적으로 학습하는 새로운 리샘플링 네트워크를 제안한다.
- 분포 모델링을 위한 변분 오토인코더(VAE, Variational Autoencoder) 도입: VAE 구조를 활용하여 [17,18] 파티클 분포를 잠재 공간(latent space)에서 모델링함으로써, 리샘플링 과정에서 파티클의 다양성을 효과적으로 보존한다.
- 적응형 하이브리드 리샘플링 전략: 유효 파티클 수(ESS, Effective Sample Size)에 따라 딥러닝 모델의 개입 정도를 조절하는 적응형 전략을 도입하여, 연산 효율성을 극대화하면서도 추정 정확도를 유지한다.

본 논문의 구성은 다음과 같다. 2장에서는 파티클 필터와 기존 리샘플링 기법의 문제점을 기술하고, 3장에서는 제안하는 파티클 트랜스포머의 구조와 학습 방법을 상세히 설명한다. 4장에서는 KITTI 데이터셋을 이용한 실험 결과 및 분석을 제시하며, 마지막으로 5장에서는 결론을 맺는다.

II. 파티클 필터

1. 파티클 필터 개요

파티클 필터는 베이지안 추론에 기반하여 비선형·비가우시안 시스템의 상태를 재귀적으로 추정하는 순차 몬테카를로(sequential Monte Carlo) 방법이다. 일반적으로 예측(prediction), 갱신(update), 리샘플링의 3단계로 구성되며, 각 시점 t 에서 N 개의 파티클 $\{x_t^i, w_t^i\}_{i=1}^N$ 로 상태 분포 $p(x_t|z_{1:t})$ 를 근사한다. 예측 단계에서는 이전 시점 $t-1$ 의 파티클 x_{t-1}^i 를 모션 모델 $f(\cdot)$ 와 제어 입력 u_t 를 이용해 현재 시점의 예측 상태로 전이시킨다. 비주얼 오도메트리(VO, Visual Odometry)에서 상태 벡터 x_t 는 차량(또는 카메라)의 6자유도 자세(pose) $[x, y, z, \phi, \theta, \psi]^T$ 로 정의될 수 있다.

$$x_t = f(x_{t-1}, u_t) + w_t \quad (1)$$

여기서 x_t 는 시점 t 에서의 상태 벡터, $f(\cdot)$ 는 비선형 모션 모델, u_t 는 제어 입력, w_t 는 모델 잡음이며 $w_t \sim N(0, Q)$ 로 모델링한다. Q 는 모델 잡음의 공분산을 의미한다.

VO의 증분 운동 Δp_t , $\Delta \theta_t$ 를 사용하는 경우, 포즈는 다음과 같이 위치와 자세로 분해하여 업데이트할 수 있다.

$$p_t = p_{t-1} + R(\theta_{t-1})\Delta p_t \quad (2)$$

$$\theta_t = \theta_{t-1} + \Delta \theta_t \quad (3)$$

여기서 p_t 는 위치(3×1), θ_t 는 오일러 각(롤(roll), 피치(pitch), 요(yaw)), $R(\cdot)$ 는 자세에 따른 회전 행렬, Δp_t 와 $\Delta \theta_t$ 는 프레임(frame) 간 상대 이동 및 회전(VO 측정으로부터 산출)을 의미한다.

갱신 단계에서는 센서 측정값 z_t 와 예측 상태를 비교하여 각 파티클의 중요도 가중치 w_t^i 를 갱신한다. 측정 모델 $h(\cdot)$ 는 상태로부터 기대 측정값을 계산하며, 측정 잡음 $v_t \sim N(0, R)$ 을 가정하면 다음과 같이 표현된다.

$$z_t = h(x_t) + v_t \quad (4)$$

여기서 z_t 는 시점 t 의 측정값, $h(\cdot)$ 는 비선형 측정 모델, R 은 측정 잡음 공분산이다. 각 파티클의 가중치는 우도(likelihood) $p(z_t|x_t^i)$ 에 비례하도록 갱신되며, 가우시안 우도를 사용하는 경우 $N(z_t; h(x_t^i), R)$ 로 계산한다.

$$w_t^i \propto p(z_t|x_t^i) \quad (5)$$

2. 리샘플링 기법의 한계

리샘플링은 낮은 가중치 파티클을 제거하고 높은 가중치 파티클을 복제하여 퇴화를 완화하는 과정이다. 체계적 리샘플링은 $O(N)$ 의 낮은 계산 복잡도로 널리 사용되지만, 고가중치 파티클의 단순 복제는 파티클 집합의 다양성을 급격히 저하시켜 샘플 빈곤화를 유발할 수 있다. 특히 동적 환경, 급회전/급가속, 저질감 구간, 또는 재지역화 상황에서는 분포가 넓게 퍼져야 하므로 다양성 손실이 추정 성능 저하로 직결된다. 정규화 리샘플링이나 마르코프 체인 몬테 카를로(MCMC, Markov Chain Monte Carlo) 기반 메트로폴리스(Metropolis) 리샘플링 등 대안이 제안되었으나, 연산 비용 증가 또는 환경 변화에 대한 적응성 부족이 여전히 남는다.

III. 파티클 트랜스포머

1. 네트워크 아키텍처(network architecture)

입력은 N 개의 파티클 상태 $x_t \in R^{N \times 6}$ 와 정규화된 가중치 $w_t \in R^{N \times 1}$ 이며, 선형 투영을 통해 256차원 임베딩으로 변환한다. 파티클 집합은 순서가 없으므로, 위치 인코딩은 시간 순서가 아니라 학습 안정성을 위한 임베딩 보조항으로만 사용한다. 인코더는 8-head 멀티헤드 어텐션과 2-layer 블록으로 구성되고, 디코더는 2-layer 크로스 어텐션으로 구성된다.

가중치 기반 멀티헤드 어텐션은 가중치 벡터 W 로부터 가중치 특징 W_{feat} 을 추출한 뒤 키(key, K)와 값(value, V)에 주입함으로써, 가중치가 큰 파티클에 더 높은 어텐션을 부여한다.

$$W_{feat} = \text{Linear}_2(\text{ReLU}(\text{Linear}_1(W))) \quad (6)$$

여기서 W_{feat} 은 가중치 특징 벡터, Linear_1 , Linear_2 는 선형 변환, ReLU 는 활성화 함수, W 는 입력 가중치 벡터이다.

$$K_{weighted} = K + W_{feat} \odot \alpha_{scale} \quad (7)$$

$$V_{weighted} = V + W_{feat} \odot \alpha_{scale} \quad (8)$$

여기서 $K_{weighted}$ 와 $V_{weighted}$ 는 가중치 주입된 키와 값, $\odot$ 는 요소별 곱, α_{scale} 은 학습 가능한 스케일링 파라미터이다. 이 구조는 단순 복제 중심의 리샘플링이 아니라, 파티클 간 관계를 학습적으로 반영한 재배치를 가능하게 한다.

VAE는 분포의 과도한 수렴을 억제하고 다양성을 유지하기 위해, 인코더 출력을 평균과 표준편차로 매핑하여 잠재분포를 학습한다. 잠재 벡터 z 는 재파라미터화 트릭으로 샘플링된다.

$$z = \mu + \sigma \odot \varepsilon, \varepsilon \sim N(0, I) \quad (9)$$

여기서 μ 는 잠재분포 평균, σ 는 표준편차, ε 는 표준 정규 분포에서 샘플링된 잡음, I 는 단위행렬이다.

디코더는 잠재 벡터 z 와 인코더 표현을 입력으로 크로스 어텐션을 수행하고, $Tanh$ 활성화와 스케일링을 통해 파티클의 상태 변화량(residual)을 예측하여 기존 파티클에 더함으로써 새로운 파티클 위치를 생성한다.

2. 적응형 하이브리드 리샘플링(adaptive hybrid strategy)

딥러닝 모듈을 매 프레임 실행하는 것은 실시간성에 부담이 되므로, ESS를 기반으로 모듈 호출 강도를 조절한다.

$$ESS = \frac{1}{\sum_{i=1}^N (w_i^t)^2} \quad (10)$$

$$ESS_{ratio} = \frac{ESS}{N} \quad (11)$$

여기서 N 은 전체 파티클 수, w_i^t 는 시점 t 에서 i 번째 파티클의 정규화된 가중치이다. ESS_{ratio} 가 낮을수록 가중치 퇴화가 심각함을 의미한다. 본 연구는 (i) $ESS_{ratio} < 0.5$ 인 경우 전체(full) 트랜스포머 모드(mode)로 70% 파티클을 갱신하고, (ii) $0.5 \leq ESS_{ratio} < 0.7$ 인 경우 부분(partial) 트랜스포머 모드로 40%만 갱신하며, (iii) $ESS_{ratio} \geq 0.7$ 인 경우 스킵(skip) 모드로 리샘플링을 생략한다. 임계값(0.5, 0.7)은 검증 기반 탐색(grid search)으로 결정한다.

3. 손실 함수(loss function)

모델 학습을 위해 위치 정확도, 자세 정확도, 그리고 잠재분포 정규화를 동시에 고려한 다목적 손실 함수를 사용한다.

$$L_{total} = 0.4L_{pos} + 0.3L_{ang} + 0.3L_{KL} \quad (12)$$

여기서 L_{pos} 는 위치 오차에 대한 평균 제곱 오차(MSE, Mean Square Error), L_{ang} 는 각도 차이를 주기성(2π)을 고려해 계산하는 자세 손실, L_{KL} 은 잠재분포 $q(z|x)$ 와 표준정규 $p(z) = N(0, I)$ 사이의 쿨백-라이블러 (KL, Kullback-Leibler) 발산이다. VAE의 KL 항은 일반적으로 다음 닫힌형 형태로 계산할 수 있다.

4. 비교 대상 리샘플링 기법

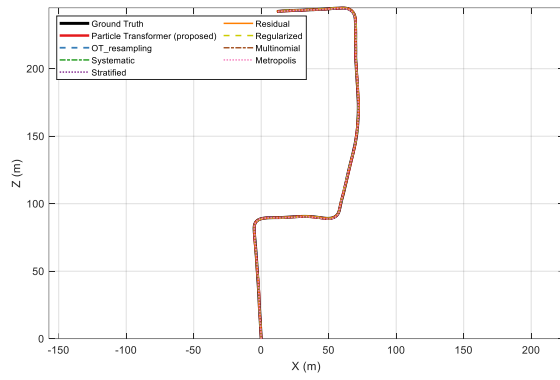
본 논문에서는 제안하는 파티클 트랜스포머의 성능 우수성을 객관적으로 입증하기 위해, 현재 널리 통용되는 7가지의 대표적인 리샘플링 기법들을 구현하여 비교 분석을 수행하였다[9,12,19]. 우선 가장 기초적인 다항 리샘플링은 각 파티클이 가진 정규화된 가중치를 해당 파티클이 선택될 확률로 간주하여 N 번의 독립적인 복원 추출을 수행하는 방식이다. 이 기법은 구현이 매우 직관적이라는 장점이 있으나, 완

전한 무작위성에 의존하기 때문에 파티클 선택의 분산(variance)이 최대가 되어 파티클 다양성을 크게 훼손할 위험이 존재한다. 이러한 분산 문제를 해결하기 위해 고안된 체계적 리샘플링은 누적 확률 분포 상에서 $[0, \frac{1}{N}]$ 구간의 난수 하나를 생성한 후, 이를 $\frac{1}{N}$ 씩 등간격으로 증가시키며 해당 구간에 위치한 파티클을 선택하는 알고리즘이다. 이 방식은 파티클 선택 과정의 무작위성을 최소화하여 분산을 효과적으로 줄이고, $O(N)$ 의 낮은 계산 복잡도를 가지기 때문에 현재 로봇 공학 및 자율주행 분야에서 가장 표준적으로 사용되고 있다. 층화 리샘플링은 체계적 리샘플링과 유사하게 분산을 줄이려는 목적을 가지나, 전체 확률 구간을 N 개의 상호 배타적인 등간격 구간으로 미리 분할한다는 점에서 차이가 있다. 각 구간 내에서 독립적으로 난수를 생성하여 파티클을 추출함으로써, 체계적 리샘플링보다 파티클 간의 독립성을 보장하면서도 다항 리샘플링보다는 낮은 분산을 유지한다. 확정적인 선택과 확률적인 선택을 결합한 잔차 리샘플링은 파티클의 가중치에 총 파티클 수를 곱한 값의 정수 부분만큼 해당 파티클을 확정적으로 복제하고, 남은 소수점 부분에 해당하는 잔여 가중치를 재정규화하여 나머지 부족한 파티클 수를 다항 리샘플링 방식으로 채우는 기법이다. 이 방법은 가중치가 높은 파티클을 확실하게 보존하면서 나머지 부분에 대해서만 확률적 기법을 적용하여 효율성을 높인다. 파티클의 다양성 확보에 중점을 둔 메트로폴리스 리샘플링은 MCMC 방법론을 차용하여, 현재 파티클 상태에서 가중치 비율에 기반한 승인-기각(accept-reject) 과정을 거쳐 파티클의 이동 여부를 결정함으로써 파티클 분포를 동적으로 갱신한다. 또한, 파티클의 중복을 방지하기 위한 정규화 리샘플링은 선택된 파티클들에 가우시안 커널과 같은 확률 밀도 함수를 적용하여 이산적인 파티클 분포를 연속적인 분포로 근사한 뒤 샘플링을 수행한다. 이를 통해 동일한 위치에 파티클이 겹치는 것을 방지하고 파티클들 사이에 인위적인 변동성을 부여하여 샘플 빈곤화 문제를 완화하고자 한다. 마지막으로 최적 수송 리샘플링은 현재의 파티클 가중치 분포를 목표로 하는 균등 분포로 변환하기 위해 필요한 운송 비용, 즉 바셔슈타인 거리(Wasserstein distance)를 최소화하는 최적의 이동 계획을 수립하여 파티클을 재배치하는 기법이다. 이는 이론적으로 원본 분포를 가장 잘 보존하는 방법 중 하나이나, 최적화 과정에서 높은 연산 비용이 요구된다는 특성이 있다 [10]. 본 논문에서는 이러한 다양한 특성을 가진 기법들과의 정량적 비교를 통해 제안 모델의 강건성과 효율성을 검증하였다.

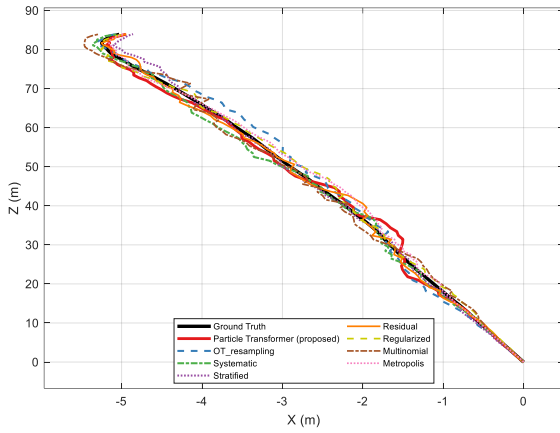
IV. 실험 및 결과

1. 실험 환경 및 데이터셋

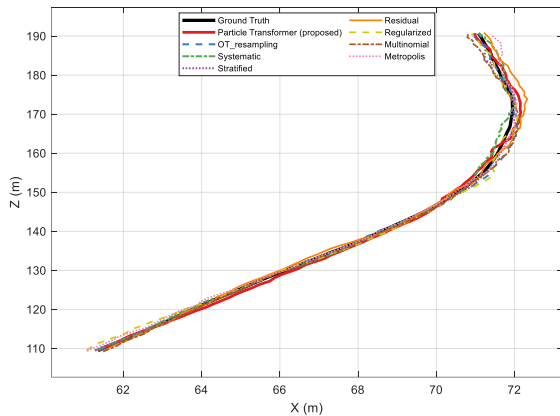
제안하는 방법의 성능을 검증하기 위해 그림 1과 같이 자율주행 벤치마크인 KITTI 오도메트리 데이터셋의 Sequence 00을 사용하였다[20,21]. 총 4,661 프레임에 대해 GPS/INS 센서 잡음을 시뮬레이션하여 학습 및 테스트 데이터를 구성하였다. 위치 잡음 표준편차 σ_{pos} 는 $[0.03, 0.03, 0.05]$ 에 잡음 레벨을 곱하고, 회전(자세) 잡음



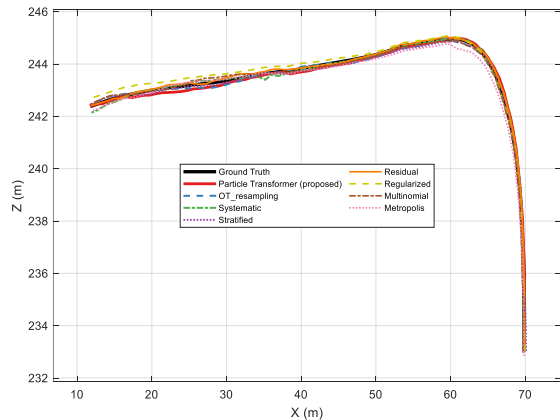
(a) Full trajectory



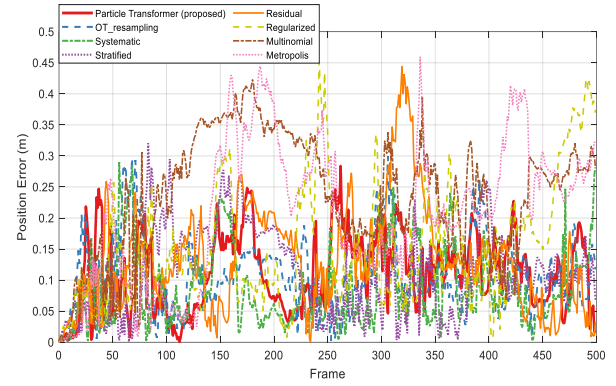
(b) Zoomed view of the first 100 frames



(c) Zoomed view of the middle 100 frames



(d) Zoomed view of the last 100 frames



(e) Euclidean position error over time

그림 1. KITTI 오도메트리 Sequence 00에서의 추정 궤적 비교(제한 기법 및 대표 비교 기법). (a) 전체 궤적. (b)-(d) 구간별 확대. (e) 프레임별 위치 오차.

Fig. 1. Estimated trajectory comparison and per-frame position error on KITTI Odometry Sequence 00: (a) Full trajectory. (b) Zoomed view of the first 100 frames. (c) Zoomed view of the middle 100 frames. (d) Zoomed view of the last 100 frames. (e) Euclidean position error over time.

표준편차 σ_{rot} 는 $[0.003, 0.003, 0.005]$ 에 잡음 레벨을 곱하여 설정하였다. 이때 잡음 레벨은 0.3에서 1.0까지 변화시키며 모델의 강건성을 평가하였다. 하드웨어 환경으로는 AMD Ryzen 9 5900X CPU와 48GB VRAM을 탑재한 NVIDIA RTX A6000 GPU를 사용하여 실험을 진행하였다.

2. 정량적 성능 분석

제한하는 파티클 트랜스포머와 기존 7가지 리샘플링 기법들의 성능을 절대 궤적 오차(ATE, Absolute Trajectory Error), 상대 포즈 오차(RPE, Relative Pose Error), 파티클 다양성, 실행 시간을 통해 비교 분석하였다.

우선 전체적인 위치 추정 정확도를 나타내는 표 1을 살펴보면, 제한하는 기법은 평균 ATE 0.1145m를 기록하여 비교군 중 가장 우수한 성능을 달성하였다. 이는 현재 자율주행 분야에서 표준적으로 사용되는 체계적 리샘플링의 0.1180 m

표 1. ATE(평균/표준편차) 및 총 실행시간 비교(KITTI Odometry Sequence 00, 4,661 프레임).

Table 1. Comparison of ATE (mean/std) and total execution time (KITTI Odometry Sequence 00, 4,661 frames).

Method	ATE Mean (m)	ATE Std.	Time (s)
Particle Transformer (proposed)	0.1145	0.0038	29.6065
OT_resampling	0.1139	0.0105	30.4097
Systematic	0.1180	0.0092	30.1564
Stratified	0.1206	0.0028	30.3302
Residual	0.1302	0.0158	30.7795
Regularized	0.1453	0.0313	30.7942
Multinomial	0.1725	0.0348	30.5765
Metropolis	0.1829	0.0285	32.9444

표 2. RPE(병진/회전), 최종 다양성, 평균 ESS 비교(KITTI Odometry Sequence 00).

Table 2. Comparison of RPE (translation/rotation), final diversity, and average ESS (KITTI Odometry Sequence 00).

Method	RPE Translation (m)	RPE Rotation (rad)	Final Diversity	Average ESS
Particle Transformer (proposed)	0.0372	0.0043	1.0000	0.9158
OT_resampling	0.0462	0.0047	0.8583	0.9314
Systematic	0.0472	0.0046	0.9033	0.9317
Stratified	0.0465	0.0050	0.7733	0.9327
Residual	0.0517	0.0054	0.7950	0.9299
Regularized	0.0504	0.0060	1.0000	0.9340
Multinomial	0.0520	0.0058	0.6100	0.9331
Metropolis	0.0429	0.0089	0.6317	0.9063

대비 약 3 %, 성능이 가장 낮은 메트로폴리스 리샘플링의 0.1829 m 대비 약 37.4 % 향상된 수치이다. 이러한 정량적 성능 차이는 그림 1의 궤적 비교 결과에서도 명확히 드러난다. 그림 1의 중간 부분(middle section)과 종료 부분(end section) 확대 그림을 보면, 기존 기법들은 곡선 구간에서 참값(GT, Ground Truth) 궤적을 이탈하는 경향을 보이는 반면, 제안 기법은 GT 궤적을 거의 완벽하게 추종함을 확인할 수 있다. 이러한 궤적 비교를 보다 정량적으로 분석하기 위해 그림 1(e)에 각 기법의 프레임별 위치 오차(유클리드 거리)를 제시하였다. 그림 1(e)에서 제안 기법(Particle Transformer)은 전 구간에 걸쳐 가장 낮은 오차를 유지하며, 특히 급격한 회전이 발생하는 구간에서도 오차가 급증하지 않고 안정적인 추정 성능을 보인다. 반면, 다항 리샘플링과 메트로폴리스 리샘플링은 시퀀스 후반부로 갈수록 오차가 누적되어 크게 증가하는 양상을 나타낸다. 이는 제안 기법의 가중치 기반 어텐션 메커니즘과 VAE를 통한 파티클 다양성 보존이 시간에 따른 오차 누적을 효과적으로 억제함을 보여준다.

국소적인 움직임 추정 성능과 파티클 필터의 건전성을 나타내는 지표는 표 2에 요약되어 있다. 제안 모델은 RPE 병진(translation) 오차 0.0372 m, 회전(rotation) 오차 0.0043 rad를 기록하며 압도적인 성능을 입증하였다. 이는 기존 기법 중 가장 양호한 OT 리샘플링(0.0462 m)과 비교해도 약 19.5%의 성능 향상을 의미하며, 가중치 기반 어텐션 메커니즘이 파티클 간의 상호작용을 학습하여 정밀한 자세 추정에 기여했음을 시사한다. 특히 표 2의 다양성(final diversity) 지표에서 제안한 기법은 1.0000의 점수를 기록하였다. 이는 다항 리샘플링(0.6100)이나 메트로폴리스 리샘플링(0.6317)과 달리, VAE를 통한 생성적 접근이 샘플 빈곤화 문제를 원천적으로 차단하고 있음을 보여준다[22].

마지막으로 성능과 연산 효율성 간의 균형을 분석하기 위해 그림 2의 ATE-실행시간 상충관계(trade-off) 그래프를 제시하였다. 그림 2에서 좌측 하단은 오차가 낮고 실행 속도가 빠른 이상적인 영역(ideal region)을 나타낸다. 제안 모델은 전체 시퀀스 처리에 29.61초(프레임당 약 6.4 ms)가 소요되어 비교군 중 가장 빠른 속도를 기록하며 해당 영역에 위치한다.

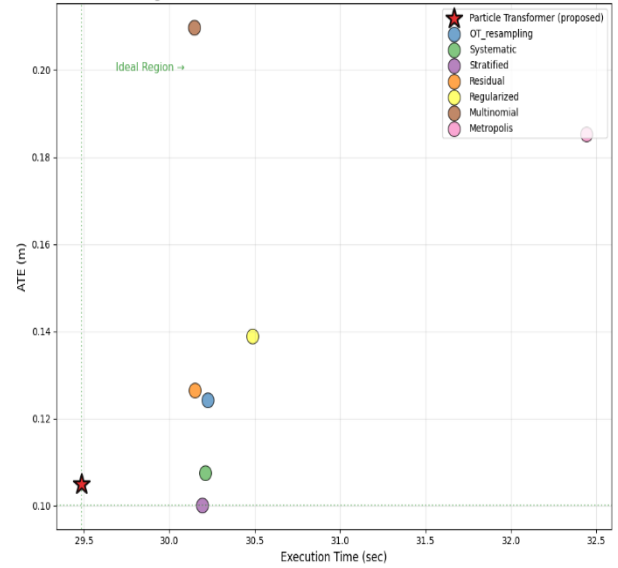


그림 2. ATE-실행시간 상충 관계 (낮을수록 우수). 각 점은 4,661 프레임 전체 평균 결과를 의미함.

Fig. 2. ATE vs. execution time trade-off (lower is better). Each point represents the average result over 4,661 frames.

표 3. 구성요소 기여도 분석.

Table 3. Component contribution analysis.

Component	Impact
Weight-Aware Attention	+8.3 % (position)
Multi-Objective Loss	+6.1 % (combined)
VAE Regularization	+4.7 % (diversity)

표 4. 손실함수 구성요소 분석.

Table 4. Loss function component analysis.

Loss Function Component	Degradation
Variance Preservation	-15.2 % (diversity)
Position Loss	-12.4 % (accuracy)
Angle Loss	-7.8 % (orientation)

이는 적응형 하이브리드 전략이 전체 프레임의 약 15 % (74 회)에서만 무거운 딥러닝 모델을 개입시키고, 나머지 구간에서는 연산 부하를 효과적으로 제거했기 때문이다.

3. 구성 요소 기여도 분석

제안하는 모델 아키텍처 내 각 모듈의 기여도를 정량적으로 파악하기 위해 핵심 모듈을 단계적으로 제거하거나 변경한 변형 모델(variant models)들을 설계하여 평가를 진행하였으며, 그 결과는 표 3과 표 4에 제시하였다.

먼저 표 3은 핵심 구성 요소의 유무에 따른 성능 변화를 보여준다. 전체 프레임워크에서 'Weight-Aware Attention' 모듈을 제거하고 일반적인 어텐션을 사용했을 경우, 위치 정확도가 약 8.3 % 하락하는 결과가 나타났다. 이는 해당 모듈이 파티클의 중요도에 따른 가중치를 명시적으로 학습하여 모델의 표현력을 강화하는 데 핵심적인 역할을 수행함을 시사한다. 또한, VAE 구조의 정규화(regularization) 기법을 제거했을 때는 파티클의 다양성 지표가 4.7 % 감소하여, 생성 모델 기반의 접근이 분포 유지에 필수적임을 확인하였다.

손실 함수의 각 항이 미치는 영향은 표 4에 세부적으로 분석되어 있다. 특히 물리적 제약 조건과 분포 특성을 반영한 'Variance Preservation(KL Loss)' 항을 제거했을 때, 다양성 지표가 15.2 %나 급감하는 현상이 발생하였다. 이는 단순한 데이터 기반 학습만으로는 파티클이 특정 모드로 붕괴(collapse)될 위험이 있음을 보여준다. 아울러 위치 오차 항(position loss)을 제외했을 때는 정확도가 12.4 % 하락하였는데, 이는 제안하는 다목적 손실 함수(multi-objective loss)가 모델의 탐색 공간을 유효한 물리적 영역 내로 제한하여 학습의 방향성을 올바르게 유도하고 있음을 입증한다.

결론적으로, 표 3과 표 4의 분석 결과는 리샘플링을 통한 데이터 품질 개선과 각 아키텍처 모듈의 상호 보완적인 결합이 제안 모델이 SOTA 수준의 성능을 달성하게 하는 핵심 요인임을 뒷받침한다.

V. 결론

본 논문에서는 순차적 중요도 리샘플링 필터의 가중치 퇴화 및 샘플 빈곤화 문제를 해결하기 위해 가중치 기반 멀티헤드 어텐션과 VAE를 결합한 파티클 트랜스포머 프레임워크를 제안하였다. KITTI 오도메트리 데이터셋을 이용한 실험 결과, 제안 기법은 체계적 리샘플링 대비 ATE 3.0 %, RPE 21.2 % 향상된 성능을 보였으며, 파티클 다양성 1.0을 유지하면서 프레임당 6.4 ms의 실시간 처리 속도를 달성하였다. 특히 ESS 기반 적응형 전략을 통해 전체 프레임의 약 15 %에서만 딥러닝 모델을 활성화하여 연산 효율성과 정확도 간의 균형을 확보하였다. 향후 연구에서는 다양한 환경 조건과 센서 구성에서의 강건성 검증 및 실제 차량 적용을 진행할 예정이다.

REFERENCES

- [1] D. Scaramuzza and F. Fraundorfer, "Visual odometry [tutorial]," *IEEE Robotics & Automation Magazine*, vol. 18, no. 4, pp. 80-92, 2011.
- [2] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1309-1332, 2017.
- [3] C. H. Kang and S. Y. Kim, "Comparative evaluation of physics-informed, hamiltonian, and lagrangian neural networks for 2D single target tracking," *Journal of Institute of Control, Robotics and Systems (in Korean)*, vol. 30, no. 3, pp. 288-293, 2024.
- [4] C. H. Kang and S. Y. Kim, "Explainable DDPG-based tracking algorithm for a steering-capable robot," *Journal of Institute of Control, Robotics and Systems (in Korean)*, vol. 30, no. 5, pp. 580-586, 2024.
- [5] J. H. Choi, C. H. Kang, and S. Y. Kim, "Advancements and performance analysis of feature detection and matching methods in visual odometry," *Journal of Institute of Control, Robotics and Systems (in Korean)*, vol. 31, no. 3, pp. 231-236, 2025.
- [6] A. J. Davison, "Real-time simultaneous localisation and mapping with a single camera," *Proc. of the 9th IEEE International Conference on Computer Vision (ICCV)*, 2003, pp. 1403-1410.
- [7] S. Wang, R. Clark, H. Wen, and N. Trigoni, "DeepVO: Towards end-to-end visual odometry with deep recurrent convolutional neural networks," *Proc. of 2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 2043-2050.
- [8] Z. Teed and J. Deng, "DROID-SLAM: Deep visual SLAM for monocular, stereo, and RGB-D cameras," *Proc. of Advances in Neural Information Processing Systems (NeurIPS)*, 2021, pp. 16558-16569.
- [9] A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*. New York, NY, USA, Springer, 2001.
- [10] P. Karkus, D. Hsu, and W. S. Lee, "Particle filter networks with application to visual localization," *Proc. of Conference on Robot Learning (CoRL)*, 2018, pp. 169-178.
- [11] R. Jonschkowski, D. Rastogi, and O. Brock, "Differentiable particle filters: End-to-end learning with algorithmic priors," *arXiv preprint arXiv:1805.11122*, 2018.
- [12] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174-188, 2002.
- [13] M. Cuturi, "Sinkhorn distances: Lightspeed computation of optimal transport," *Proc. of Advances in Neural Information Processing Systems (NeurIPS)*, 2013.
- [14] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *Proc. of Advances in Neural Information Processing Systems (NeurIPS)*, 2017.
- [15] A. Dosovitskiy, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.
- [16] C. H. Kang and S. Y. Kim, "Robust visual odometry via attention-based adaptive particle resampling," *Proc. of 2025 ICROS-KROS Honam-Jeju Regional Conference (in Korean)*, pp. 121-122, Kunsan, Korea, Dec. 2025.
- [17] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," *arXiv preprint arXiv:1312.6114*, 2013.
- [18] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner, "beta-VAE: Learning basic visual concepts with a constrained variational framework," *Proc. of International Conference on Learning Representations (ICLR)*, 2017.
- [19] T. Li, M. Bolic, and P. M. Djuric, "Resampling methods for particle filtering: Classification, implementation, and strategies," *IEEE Signal Processing Magazine*, vol. 32, no. 3, pp. 70-86, 2015.
- [20] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," *Proc. of 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 3354-3361.
- [21] J. Graeter, A. Wilczynski, and M. Lauer, "LIMO: Lidar-monocular visual odometry," *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 7872-7879.
- [22] H. Shin, J. K. Lee, J. Kim, and J. Kim, "Continual learning with deep generative replay," *Proc. of Advances in Neural Information Processing Systems (NeurIPS)*, 2017.



강 창 호

2009년 세종대학교 기계항공우주공학과 졸업. 2016년 서울대학교 공과대학 기계항공공학부 대학원 졸업(공학박사). 2016년~2018년 BK21 플러스 서울대학교 융합지식기반 창조형 기계항공인재양성사업단 박사후 연구원. 2018년~ 2019년 고려대학교 공학연구소 연구교수. 2019년~2024년 국립금오공과대학교 기계시스템공학과(항공기계전자융합전공) 조교수. 2024년~현재 세종대학교 AI로봇학과 조교수. 관심분야는 항법 시스템, 딥러닝, 디지털 신호 처리, 비선형 필터링, 다중 목표물 추적 기법.



김 선 영

1999년 국민대학교 전자공학과 졸업. 2015년 서울대학교 공과대학 기계항공공학부 대학원 졸업(공학석사). 2019년 동 대학원 졸업(공학박사). 2019년~2020년 세종대학교 지능기전공학부 박사후 연구원. 2020년~2023년 국립군산대학교 기계공학부 조교수. 2023년~현재 국립군산대학교 기계공학부 부교수. 관심분야는 항법 시스템, 필터링, 자율주행, 딥러닝, 다중 센서 융합, 다중 목표물 추적, 목표물 검출 및 식별, 고장 검출 및 진단.