

# 차량-기지국 간 통신에서 심층 강화 학습 기반 빔 추적 알고리즘과 실험

박현우, 강정완, 김선우

한양대학교 전자컴퓨터통신공학과

{stark95, rkdwjddhks77, remero}@hanyang.ac.kr

## Experimental Results of Deep Reinforcement Learning Based Beam Tracking Algorithm for V2I Communications

Hyunwoo Park, Jeongwan Kang, and Sunwoo Kim

Department of Electronics and Computer Engineering, Hanyang University

### 요약

본 논문에서는 밀리미터파 차량-기지국 간 통신 상향링크 SIMO 채널에서 심층 강화 학습 기반 빔 추적 알고리즘을 제안한다. 밀리미터파 V2I 통신은 고속 이동성을 가진 차량 환경을 고려하며, 이 때, 차량-기지국 간 통신 링크를 유지하기 위한 실시간 빔 추적이 필수적이다. 하지만 기존 필터 기반 빔 추적 알고리즘은 높은 계산 복잡도로 인해 고속 이동 차량 환경에 적합하지 않다. 이를 해결하기 위해 매시간 수신 신호를 관측, 심층 강화 학습을 통해 실시간으로 차량 이동성을 따라 빔을 추적하는 알고리즘을 제안하며 시뮬레이션을 통해 추적 성능을 비교 분석한다.

### I. 서론

다중 안테나 기술은 5G 요구사항을 충족하는 기술 중 하나이며, 고 지향성을 위해 안테나 수가 증가하는 추세이다[1]. 안테나 수가 증가하면 빔 폭이 좁아지며, 이는 이동 통신 시스템에서 더욱 정확하고, 빈번한 빔 정렬을 요구한다. 기존 필터 기반 빔 추적 알고리즘은 차량 이동성을 고려하나[2], 높은 계산 복잡도로 인해 실시간 빔 추적에 적합하지 않다.

본 논문에서는 차량-기지국(vehicle-to-infrastructure, V2I) 통신 상향링크 SIMO(single-input-multiple-output) 채널에서 심층 강화 학습 기반 실시간 빔 추적 알고리즘을 제안한다. 특히 매우 빠른 액션이 가능하여 실시간 빔 추적에 적합한 DQN(deep Q-network)[3]을 사용하는 빔 추적 알고리즘을 설계, 성능을 확인한다.

### II. V2I 통신 심층 강화 학습 기반 빔 추적 알고리즘

본 논문에서는 2 차원 평면에서 차량의 움직임을 고려한다. 시간  $t_k$ 은  $t_k = t_0 + k\Delta t$ 이며,  $k$ 은 타임 슬롯 인덱스,  $\Delta t$ 는 타임 슬롯 간격이다. 이동성 상태 벡터  $\mathbf{m}_k$ 는 다음과 같다.

$$\mathbf{m}_k = [x_k, \dot{x}_k, \ddot{x}_k, y_k, \dot{y}_k, \ddot{y}_k]^T, \quad (1)$$

$x_k, \dot{x}_k, \ddot{x}_k, y_k, \dot{y}_k, \ddot{y}_k$ 는 각각  $x - y$  평면에서 차량의 위치, 속도 및 가속도를 나타낸다. 차량의 이동성 모델은 명령 프로세스  $\mathbf{u}_k = [u_{x,k}, u_{y,k}]^T$  및 가속도 잡음  $\mathbf{v}_k = [v_{x,k}, v_{y,k}]^T$ 으로 구성된다.  $\mathbf{u}_k$ 는 마르코프 체인으로 결정되며, 가속도 잡음  $\mathbf{v}_k$ 는 평균 0, 표준편차  $\sigma_v$ 을 가진다. 이때 이동성 모델은 다음과 같다[4].

$$\mathbf{m}_{k+1} = \mathbf{A}(\Delta t, \beta)\mathbf{m}_k + \mathbf{B}_u(\Delta t)\mathbf{u}_k + \mathbf{B}_v(\Delta t)\mathbf{v}_k, \quad (2)$$

$$\mathbf{A}(\Delta t, \beta) = \begin{bmatrix} \tilde{\mathbf{A}} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & \tilde{\mathbf{A}} \end{bmatrix}, \mathbf{B}_i(\Delta t) = \begin{bmatrix} \tilde{\mathbf{B}}_i & \mathbf{0}_{3 \times 1} \\ \mathbf{0}_{3 \times 1} & \tilde{\mathbf{B}}_i \end{bmatrix}, \quad (3)$$

$$\tilde{\mathbf{A}} = \begin{bmatrix} 1 & \Delta t & \frac{\Delta t^2}{2} \\ 0 & 1 & \Delta t \\ 0 & 0 & \beta \end{bmatrix}, \tilde{\mathbf{B}}_u = \begin{bmatrix} \frac{\Delta t^2}{2} \\ \Delta t \\ 0 \end{bmatrix}, \tilde{\mathbf{B}}_v = \begin{bmatrix} \frac{\Delta t^2}{2} \\ \Delta t \\ 1 \end{bmatrix}, \quad (4)$$

$\beta$ 는 연속적인 타임 슬롯 간의 가속 상관도이며, 인덱스  $i$ 는  $u$ 또는  $v$ 이다.

채널 상태 벡터  $\mathbf{x}_k$ 는 다음과 같다.

$$\mathbf{x}_k = [\text{Real}(\alpha_k), \text{Imag}(\alpha_k), \theta_k]^T, \quad (5)$$

$\alpha_k$ 과  $\theta_k$ 은 각각 채널 이득과 도래각(angle of arrival, AoA)이며,  $\alpha_k$ 은 아래 식과 같이 1 차원 마르코프 모델을 따른다.

$$\alpha_{k+1} = \rho\alpha_k + \zeta_k, \quad (6)$$

$\rho$ 는 상관 계수이며,  $\zeta_k$ 는  $\zeta_k \sim \mathcal{CN}(0, 1 - \rho^2)$ 을 따른다. 도래각  $\theta_k$ 는 이동성 모델에서 차량의 기지국에 대한 상대적 각도로 표현되며 식은 다음과 같다.

$$\theta_k = \tan^{-1}\left(\frac{y_k}{x_k}\right). \quad (7)$$

본 논문에서는 단일 안테나를 가진 차량과  $N$ 개의 선형 배열 안테나(uniform linear array, ULA)를 가진 기지국 간 단일 경로 SIMO 채널을 고려하며 기지국 빔포밍 벡터  $\mathbf{w}_k$ 와 채널 모델  $\mathbf{h}_k$ 은 다음과 같다.

$$\mathbf{w}_k = \mathbf{a}(\hat{\theta}_k), \quad (8)$$

$$\mathbf{h}_k = \alpha_k \mathbf{a}(\theta_k), \quad (9)$$

$\hat{\theta}_k$ 는 기지국 빔 조향 각도이며, 어레이 벡터  $\mathbf{a}(\theta_k)$ 은 다음과 같다.

$$\mathbf{a}(\theta_k) = \frac{1}{\sqrt{N}} [1, e^{j\pi \cos \theta_k}, \dots, e^{j\pi(N-1) \cos \theta_k}]^T. \quad (10)$$

심층 강화 학습 중 하나인 DQN은 에이전트(agent)가 매 시간 상태(state)  $\mathbf{s}_k$ 를 관측해 기대 보상(cumulative future reward)을 최대화하는 동작(action)  $\mathbf{a}_k$ 를 수행하고, 보상(reward)  $\lambda_k$ 를 받아 신경망(network)을 학습한다. DQN의 액션 이전/이후 관측 모델(measurement model)  $r_k^{\text{bf}}, r_k^{\text{af}}$ 은 다음과 같다.

$$r_k^{\text{bf}} = \mathbf{w}_{k-1}^H \mathbf{h}_k + n_k^{\text{bf}}, \quad (11)$$

$$r_k^{\text{af}} = \mathbf{w}_k^H \mathbf{h}_k + n_k^{\text{af}}, \quad (12)$$

노이즈  $n_k^{bf}, n_k^{af}$ 는  $n_k^{bf}, n_k^{af} \sim \mathcal{CN}(0, \sigma_n^2)$ 을 따르며, DQN의  $\mathbf{s}_k$ 는 아래와 같다.

$$\mathbf{s}_k = \left[ \text{real}\left(\frac{r_k^{bf}}{r_k^{af}}\right), \text{imag}\left(\frac{r_k^{bf}}{r_k^{af}}\right), \hat{\theta}_{k-1} \right]^T. \quad (13)$$

DQN의  $\mathbf{a}_k$ 는 집합  $\{0, 1, 2\}$ 원소 중 하나이며, 다음의 동작을 취한다.

$$\hat{\theta}_k = \hat{\theta}_{k-1} + (a_k - 1)\Delta\hat{\theta}, \quad (14)$$

$\Delta\hat{\theta}$ 는 액션의 각도 변화량이다. DQN의  $\lambda_k$ 는 다음과 같이 정의한다.

$$\lambda_k = \begin{cases} 1, & \text{if } |r_k^{af}/r_k^{bf}| > c_u \\ -1, & \text{if } |r_k^{af}/r_k^{bf}| < c_l \\ 0, & \text{otherwise.} \end{cases} \quad (15)$$

### III. 시뮬레이션 결과

본 논문에서는 신호 대 잡음 비 SNR 및 기지국 안테나 개수에 따른 빔 추적 성능을 비교 분석한다. 송신 전력 1 일 때, SNR [dB]은 다음과 같다.

$$\text{SNR} = 10 \log_{10} \left( \frac{1}{\sigma_n^2} \right). \quad (16)$$

차량은 최대 20m/s의 속도를 가지며, 매 사건(episode) x 좌표 [-100m, 100m], y 좌표 [0m, 200m]범위 내 무작위 위치에서 이동성 모델을 따라 움직인다. 기지국의 좌표는 (0m, 0m)이며, 차량은 기지국과 최소 10m 거리를 유지한다. 차량의 최대 속도와 차량-기지국 간 최소 거리를 이용해 타임 슬롯 간 도래각 변화량  $\Delta\theta$ 의 상한값은 다음과 같이 구할 수 있다.

$$\Delta\theta \leq \frac{20[\text{m/s}]}{10[\text{m}]} \cdot 10[\text{ms}] = 0.02 [\text{rad}]. \quad (17)$$

매 사건 초기 빔 정렬이 되어 있다고 가정하고 빔 추적을 수행한다. 동작 각도 변화량  $\Delta\theta$ 은 최대 도래각 변화량보다 크거나 같다. 200 번의 사건 동안 타임 슬롯에 따라 실제 도래각과 추적한 도래각의 MSE(mean squared error)를 측정하여 성능을 비교한다. 시뮬레이션 파라미터는 표 1과 같다.

그림 1에서  $N$ 을 32로 고정하고, 다양한 SNR에서 타임 슬롯에 따른 빔 추적 성능을 비교한다. SNR이 증가하면, 노이즈의 영향이 줄어들어 보다 정확한 빔 추적이 가능하다. 그림 2에서 SNR을 30dB로 고정하고, 다양한  $N$ 에서 타임 슬롯에 따른 빔 추적 성능을 비교한다.  $N$ 이 증가하면 빔 폭이 줄어들고, 지향하는 방향에 대한 안테나 이득이 커지며,  $N$ 이 증가함에 따라 추적 성능이 향상된다.

표 1. 시뮬레이션 및 DQN 파라미터

Parameter	Value
기지국 안테나 원소 개수 $N$	8, 16, 32
타임 슬롯 간격 $\Delta t$	10 [ms]
채널 이득 상관 계수 $\rho$	0.995
타임 슬롯 가속도 상관도 $\beta$	0.9
동작 도래각 변화량 $\Delta\hat{\theta}$	0.02 [rad]
상한 임계 값 $c_u$	1.01
하한 임계 값 $c_l$	0.99
노이즈 분산 $\sigma_n^2$	$10^{-1}, 10^{-2}, 10^{-3}$

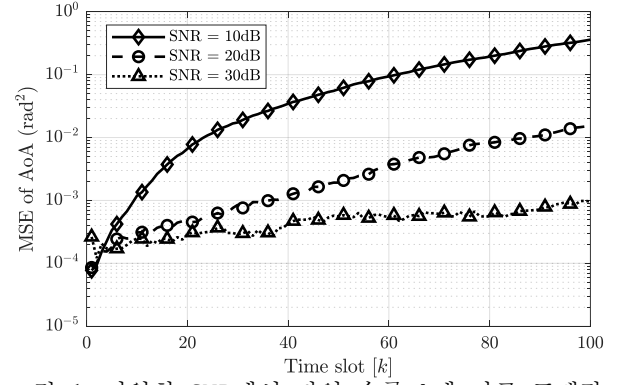


그림 1. 다양한 SNR에서 타임 슬롯  $k$ 에 따른 도래각 (AoA) 추적 성능 비교 ( $N = 32$ )

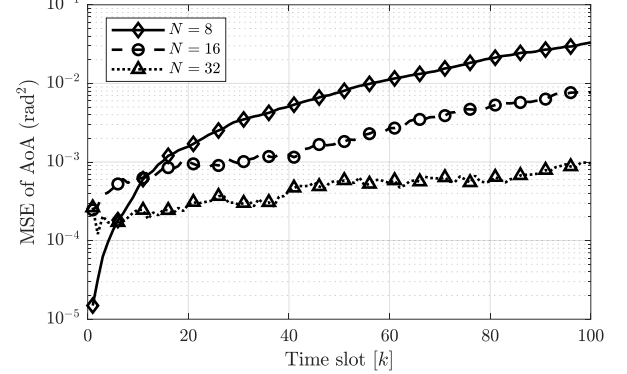


그림 2. 다양한  $N$ 에서 타임 슬롯  $k$ 에 따른 도래각 (AoA) 추적 성능 비교 (SNR = 30dB)

### IV. 결론

본 논문에서는 차량-기지국 간 상향링크 SIMO 채널에서 심층 강화 학습 기반 실시간 빔 추적 알고리즘을 제안한다. 기존의 필터 기반 빔 추적에서 발생하는 연산 시간 문제를 해결하기 위해 DQN을 도입한다. 먼저, 다양한 SNR에 대한 DQN의 추적 성능을 비교하고, 두번째로 다양한  $N$ 에 대한 추적 성능을 비교한다. 시뮬레이션 결과 SNR이 클수록, 그리고  $N$ 이 증가할수록 빔 추적 성능이 향상됨을 확인하였다.

### ACKNOWLEDGMENT

이 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (2016-0-00208, 차세대 5G V2X 서비스 실현을 위한 정밀 측위탐색 연계 고효율 다중안테나 정보전송 및 네트워크 기술 연구)

### 참고 문헌

- [1] A. Yadav *et al.*, "Energy and Traffic Aware Full-Duplex Communications for 5G Systems," *IEEE Access*, vol. 5, pp. 11278-11290, May. 2017.
- [2] J. Lim *et al.*, "Beam Tracking under Highly Nonlinear Mobile Millimeter-wave Channel," *IEEE Commun. Lett.*, vol. 23, no. 3, pp. 450-453, Mar. 2019.
- [3] V. Mnih *et al.*, "Human-level Control through Deep Reinforcement Learning," *Nature*, vol. 518, pp. 529-533, Feb. 2015.
- [4] T. Liu *et al.*, "Mobility Modeling, Location Tracking, and Trajectory Prediction in Wireless ATM Networks," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 6, pp. 922-936, Aug. 1998.