

밀리미터파 및 테라헤르츠 무선 통신을 위한 컴퓨터 비전 기반 빔포밍

김 선 우*, 김 재 우*, 이 종 만*, 심 병 호^o

Computer Vision-Aided Beamforming for Millimeter and Terahertz Communications

Sunwoo Kim*, Jaewoo Kim*, Jongman Lee*, Byonghyo Shim^o

요 약

최근 센싱과 컴퓨터 비전 (Computer Vision, CV) 기술의 발전으로 인해 딥러닝 (Deep Learning, DL) 기반 컴퓨터 비전 기술이 밀리미터파 및 테라헤르츠 빔포밍에 적용될 수 있게 되었다. 컴퓨터 비전 기반 빔포밍에서는 사전 훈련된 물체 탐지기 (pre-trained object detector)를 사용하여 무선 객체의 채널 파라미터 정보 (거리 및 각도)를 획득하고, 기지국이 해당 위치로 빔을 전송하는 방식으로 이루어진다. 본 논문에서는 이러한 이득을 극대화하기 위해 무선 객체를 탐지하는데 특화된 물체 탐지기를 사용하여 빔을 생성하는 새로운 빔포밍 기법을 제안한다. 물체 탐지기의 학습을 위해 휴대폰 및 노트북을 가진 사람들의 RGB 및 심도 이미지로 구성된 Vision Objects for Millimeter and Terahertz Communications (VOMTC)라는 거대한 비전 데이터셋을 수집하였다. VOMTC 테스트 셋 (test set)을 활용한 빔포밍 실험 결과를 통해 제안하는 기법이 기존 빔포밍 기법 대비 데이터 전송률을 크게 향상시킨다는 것을 확인하였다.

Key Words : Dataset, Deep Learning, Sensing, Computer Vision, 6G Wireless Communications

ABSTRACT

Recent progress in sensing and computer vision has enabled the application of deep learning (DL)-based CV technology in millimeter and terahertz beamforming. In CV-based beamforming, a pre-trained object detector is used to acquire the channel parameter information (distance and angle) of the wireless object, using which the BS can generate a directional beam heading toward the wireless object. In this paper, we propose a novel beamforming technique that exploits an object detector tailored for identifying wireless objects to maximize the beamforming gain. To develop such detector, we collected a massive vision dataset called Vision Objects for Millimeter and Terahertz Communications (VOMTC), which consists of RGB and depth images of people holding cell phones and laptops. Through beamforming experiments using the VOMTC test set, we show that the proposed technique outperforms conventional beamforming techniques in terms of data rate.

* This work was supported by Korea Research Institute for Defense Technology (KRIT) - Grant funded by Defense Acquisition Program Administration (DAPA) (KRIT-CT-22-036).

♦ First Author : Seoul National University Department of Electrical & Computer Engineering, swkim@islab.snu.ac.kr, 학생회원

° Corresponding Author : Seoul National University Department of Electrical & Computer Engineering, bshim@snu.ac.kr, 종신회원

* Co-Authors: Hanwha Systems, jaewooz@hanwha.com, jongman0319.lee@hanwha.com

논문번호 : 202407-139-A-RN, Received July 10, 2024; Revised August 10, 2024; Accepted August 21, 2024

1. 서론

최근 밀리미터파 (mmWave) 및 테라헤르츠 (THz) 대역 통신은 6G를 위한 데이터 요구가 많고 지연에 민감한 다양한 응용 프로그램을 지원하는 데 핵심적인 기술로 주목받고 있다^[1]. mmWave 및 THz 대역 (24 GHz ~10 THz)의 풍부한 주파수 스펙트럼 자원을 활용함으로써, mmWave 및 THz 통신은 디지털 트윈, 메타버스, 고품질 홀로그램과 같은 진정으로 몰입형 서비스를 실현하는 데 기여할 수 있다^[2].

이러한 초고주파 통신의 주요한 단점은 전파, 반사, 확산 산란 및 대기 흡수 손실로 인한 신호 전력의 심각한 감쇠이다. 이러한 심각한 경로 손실을 보상하기 위해 대용량 다중 입력 다중 출력(massive multiple-input multiple-output, massive MIMO) 시스템을 사용하는 빔포밍 (beamforming) 기술이 필요하다^[3]. 빔포밍 이득을 극대화하기 위해서는 빔포밍 벡터가 신호 전파 경로의 배열 스티어링 (array steering) 벡터와 정렬되어야 하고, 이를 위해 기지국(Base Station, BS)은 핸드폰과 같은 무선 객체의 정확한 다운링크 채널 정보를 확보해야 한다^[4].

종래의 5G NR 시스템에서는 채널 정보 획득을 위해 빔 코드북이라고 불리는 아날로그 빔 코드워드 집합을 사용해왔다^[5]. 이러한 코드북 기반의 5G NR 빔포밍에서는 다양한 문제점들이 존재한다. 우선, 미리 정의된 빔 방향과 실제 방향 간의 불일치로 인해 5G NR 빔포밍 이득의 감소가 발생한다. 또한, mmWave/THz 채널 코히런스 타임 (coherence time) (예: 단말이 30km/h 속도로 이동할 때 9ms)이 짧기 때문에, 단말이 약간만 움직여도 빔 방향이 불일치될 수 있으며, 결과적으로 빔포밍 이득이 줄어들게 된다^[6].

이에 대한 대안으로써, 컴퓨터 비전 (Computer Vision, CV) 기반 빔포밍 기술이 제안되었다^[7]. 이 기술에서는 기지국 주변의 무선 통신 환경을 카메라 센싱 정보로 캡처하여 무선 객체의 정확한 위치 정보를 얻기 위한 딥러닝 (Deep Learning, DL) 기반의 물체 탐지기 (object detector)의 입력으로 사용한다. 또한, 라이다 센서를 RGB 센서와 공동으로 활용하면 거리 정보와 더불어 2차원 이미지 상에서의 객체 좌표를 얻을 수 있고, 나아가 무선 객체의 3차원 좌표를 획득할 수 있다. 3차원 좌표는 무선 채널 파라미터 정보인 무선 객체의 거리 및 방위각과 고도각으로도 표현되므로, 이를 통해 얻어지는 방향으로 빔을 전송하면 빔포밍 이득과 데이터 전송률을 크게 향상시킬 수 있다.

본 논문에서는 이러한 이득을 극대화하기 위해 (즉, 이미지에서 무선 객체를 보다 정확하게 탐지하기 위해) 새로운 CV기반 빔 관리 기법인 VOMTC-aided beam management (VBM)을 제안한다. 고양이, 강아지, 의자와 같은 다양한 물체를 탐지하기 위해 사용되는 일반적인 물체 탐지기 (예: YOLO^[8], EfficientDet^[9]) 기반 빔포밍 기법과는 다르게, VBM은 무선 객체만을 탐지하는데 특화된 물체 탐지기를 활용한다. 이러한 탐지기를 얻어내기 위해, EfficientDet을 기준으로 삼고, 수집한 데이터셋인 Vision Objects for Millimeter and Terahertz Communications (VOMTC)를 사용하여 정밀 튜닝 (Fine-tuning, FT)을 수행하였다.

본 연구에서 수집한 VOMTC 데이터셋은 총 20,232 장으로 이루어져 있으며 휴대폰 및 노트북을 가진 사람들의 RGB 픽셀 및 심도(depth) 이미지 쌍으로 구성되어 있다. 구체적으로는 그림 1의 예시처럼 1,000개 이상의 다양한 장소에서 사람들의 이미지 데이터를 수집하였다(<https://github.com/islab-github/VOMTC> 참조). 이미지에서 무선 객체를 보다 정확하게 탐지하는 맞춤형 물체 탐지기를 사용함으로써, VBM은 무선 객체로 향하는 방향성 빔을 전송할 확률을 향상시킬 수 있었고, 실험을 통해 기존 5G 빔 관리 기법 대비 월등히 높은 데이터 전송률을 얻을 수 있음을 확인하였다.

본 논문의 구성은 다음과 같다. 서론에 이어 II장에서 시스템 모형 및 컴퓨터 비전 기반 빔포밍을 설명한다. III장과 IV장에서는 각각 VOMTC 데이터셋 획득과 VOMTC로 학습된 빔 관리 기법을 소개한다. V장에서는 모의실험 결과를 설명하고, V장에서 결론을 지으며 논문을 마무리한다.



그림 1. 본 연구에서 획득한 VOMTC 데이터셋 예시
Fig. 1. Snapshots of VOMTC dataset

II. 밀리미터파 및 테라헤르츠 빔포밍

2.1 밀리미터파 및 테라헤르츠 시스템 모형

본 논문에서는 M개의 균일 평면 배열 안테나 (Uniform Planar Array, UPA)를 가지는 기지국과 단일 안테나를 가지는 K개의 단말로 이루어진 다운링크 다중 입력 단일 출력 (MISO) 시스템을 고려한다. mmWave/THz 신호의 낮은 침투력과 강한 직진성으로 인해 신호 전송은 주로 시야선(Line of Sight, LoS) 링크에 의해 지배된다. 따라서 기지국에서 k번째 단말로의 다운링크 채널 벡터는

$$\mathbf{h}_k = \sqrt{\beta_k} \alpha_k \mathbf{a}_M(\theta_k, \phi_k) \quad (1)$$

이다. 이때, β_k 는 경로 손실 등을 포함한 대규모 페이딩 계수이고, $\alpha_k \sim CN(0,1)$ 는 소규모 페이딩 계수이다. θ_k 와 ϕ_k 는 각각 k번째 단말의 방위각과 고도각이다. 그리고, UPA 안테나 시스템의 응답인 스티어링 벡터 $\mathbf{a}_M(\theta_k, \phi_k)$ 는 다음과 같다.

$$\begin{aligned} \mathbf{a}_M(\theta_k, \phi_k) &= \frac{1}{\sqrt{M}} [1 \dots e^{-j\pi(M_x-1)\cos(\theta_k)\sin(\phi_k)}]^T \\ &\otimes [1 \dots e^{-j\pi(M_y-1)\sin(\theta_k)\sin(\phi_k)}]^T \end{aligned} \quad (2)$$

이때, \otimes 는 크로네커 곱이다.

본 논문에서는 기지국이 M개의 아날로그 위상 변조기(analog phase shifter)를 사용한다는 일반적인 가정을 한다. 기지국의 송신 전력 제약으로 인해, 아날로그 프리코더 $\mathbf{f}_k = \mathbf{a}_M(\theta_k, \phi_k)$ 는 $\|\mathbf{f}_k\|_2^2 = 1$ 와 같이 정규화된다. 이때, 단말 k의 수신 신호는

$$y_k = \sqrt{P_k} \mathbf{h}_k^H \mathbf{f}_k s_k + n_k \quad (3)$$

이다. 이때, s_k 는 k번째 단말을 위한 데이터 심볼이고, P_k 는 송신 신호의 전력이며, n_k 는 분산이 σ_n^2 인 가우시안 잡음이다. 따라서, k번째 단말의 달성 가능한 데이터 전송률은 다음과 같이 표현된다.

$$R_k = \log_2(1 + P_k |\mathbf{h}_k^H \mathbf{f}_k|^2 / \sigma_n^2) \quad (4)$$

또한, K개의 단말의 총 데이터 전송률은 $R_{total} = \sum_{k=1}^K R_k$ 이다.

R_{total} 은 다중 사용자 간섭(inter-user interference)의 영향을 받을 수 있다. 그러나 안테나 수가 충분히 많은 경우, 단말들의 채널 벡터가 점근적으로 서로 직교하게 된다^[10]. 이는 수백 또는 수천 개의 안테나를 사용하는 차세대 6G 시스템(예: 지능형 반사판, 홀로그래프 MIMO)에서도 유효하며, 이로 인해 다른 단말들의 위치를 향한 빔들에 의해 발생하는 다중 사용자 간섭이 무시할 수 있을 정도로 작아질 수 있음을 의미한다. 따라서 본 논문에서는 단말 간의 간섭이 거의 없다고 가정 한 OMA(Orthogonal Multiple Access) 시스템을 고려하였으며, 식 (3)과 식 (4)는 이러한 OMA 설정 하에서 유도되었다.

다중 사용자 간섭이 있을 경우, 기존의 간섭 관리 기법들을 추가적으로 사용할 수 있다. 예를 들어, NOMA(Non-Orthogonal Multiple Access) 시스템에서는 여러 단말의 신호가 중첩되면서 간섭이 발생하는 데, 이를 효과적으로 제어하기 위해 SIC(successive interference cancellation)와 MPA(message passing algorithm)를 사용할 수 있다. 또한, LDS(low-density signature) 시퀀스를 사용하여 각 단말의 코드워드를 희소하게 표현함으로써, NOMA 시스템에서의 다중 사용자 간섭을 줄이는 데 기여할 수 있다^[11].

2.2 컴퓨터 비전 기반 빔포밍

앞서 언급한 바와 같이 지금까지의 밀리미터파 및 테라헤르츠 빔포밍은 여러 가지 문제점들이 존재한다. 5G NR 빔포밍의 잘 알려진 단점 중 하나는 사전에 정의된 이산적인 빔 방향과 실제 방향 간의 불일치이다. 예를 들어, 1024개의 빔이 사용되는 5G NR 시스템에서 각 빔은 평균적으로 8 제곱도의 원형 섹터를 커버하므로, 최악의 경우 방위각과 고도각에서 4도 가량의 각도 오차가 발생한다. 이는 빔포밍 이득의 20% 이상 감소로 이어진다^[12]. 또 다른 단점은 빔 스위프 (beam sweeping) 및 빔 세밀화 (beam refinement) 작업으로 인한 지연이다. 빔 세밀화 과정에서는 4개의 채널 상태 정보 기준 신호 (channel state information-reference signal, CSI-RS) 빔을 전송하는 데 30ms가 소요된다. mmWave/THz 채널의 짧은 코히런스 타임 (예: 이동 중인 무선 객체의 속도가 30km/h일 때 9ms) 때문에, 객체의 작은 이동으로도 빔 방향이 완전히 불일치할 수 있으며, 결과적으로 빔포밍 이득이 감소할 수 있다^[12].

이를 극복하기 위해 컴퓨터 비전 기반 빔포밍이 사용될 수 있다. 이 방식의 주요 이점은 코드북에 의존하지 않고 획득한 센싱 정보로부터 바로 정확한 빔 방향을 알아낼 수 있으므로, 기존 5G NR의 복잡한 피드백 및

빔 트레이닝 과정을 생각할 수 있다는 것이다¹¹⁾. 컴퓨터 비전 기반 빔포밍은 RGB-d 카메라 및 라이다와 같은 다양한 이미징 센서들을 활용하여 단말의 위치를 식별하고, 그 후 식별된 위치로 빔 $a_M(\theta_k, \phi_k)$ 을 생성한다.

구체적으로, 컴퓨터 비전 기반 빔포밍은 먼저 물체 탐지기를 사용하여 이미지에서 무선 객체의 2D위치 (x, y)를 찾는다. 본 연구에서는 수집한 VOMTC 데이터셋을 사용하여 무선 객체만을 탐지하는데 특화된 물체 탐지기를 개발하고자 한다. 2D위치 (x, y)가 연어지면, 무선 객체의 방위각과 고도각을 찾을 수 있다. 구체적으로, 컴퓨터 비전 기반 빔포밍은 심도 카메라 (depth camera)를 사용하여 심도 이미지를 캡처한다. 이 센싱 정보는 각 픽셀에 대한 지점까지의 거리를 포함한다. 얻어진 깊이 정보 r 과 무선 객체의 2D 위치를 사용하여 3D 좌표 (x, y, z)를 쉽게 찾을 수 있으며 ($z = \sqrt{r^2 - (x^2 + y^2)}$), $\theta = \tan^{-1}(\sqrt{x^2 + y^2}/z)$ 와 $\phi = \tan^{-1}(y/x)$ 를 사용하여 무선 객체의 방위각과 고도각으로 변환할 수 있다 (그림 2 참조).

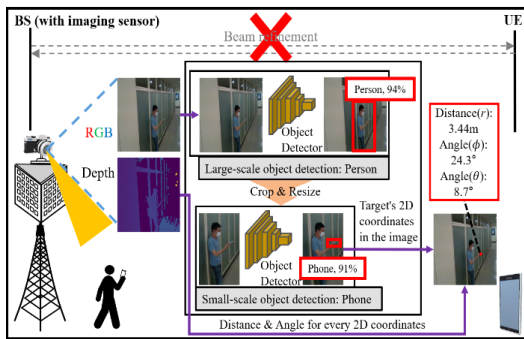


그림 2. 본 연구에서 획득한 데이터셋을 무선 객체 식별 및 위치 추정에 활용하는 예시
Fig. 2. Dataset use case: phone identification and localization

III. VOMTC 데이터셋 획득

무선 객체를 탐지하는데 특화된 물체 탐지기를 학습시키기 위해서는 데이터셋이 필요하다. 이를 위해 본 연구에서는 기지국에 부착된 RGB-depth 카메라의 시점에서 촬영된 이미지들을 수집하였다. 이렇게 수집된 데이터셋은 Vision Objects for Millimeter and Terahertz Communications (VOMTC)라고 명명하는 20,232쌍의 RGB 및 심도 이미지로 구성된다. 이 데이터셋은 인텔 리얼센스 L515 RGB-d 카메라로 수집되었

다. VOMTC의 각 이미지에는 사람, 휴대폰, 노트북과 같은 클래스에 대한 라벨링이 되어 있으며 물체들이 이미지에서 차지하는 영역을 나타내는 바운딩 박스가 표시되어 있다 (그림 3 참조). 그 후, VOMTC는 각각 17,299, 1,400, 및 1,493개의 RGB-심도 이미지 쌍을 포함하는 훈련, 검증, 및 테스트 세트로 나누어졌다.

주로 큰 규모의 물체 (예: 자동차, 버스, 트럭)를 포함하고 기존의 무선 데이터셋과는 달리, VOMTC는 작은 규모의 무선 물체(무선 객체 및 노트북)에 대한 라벨(즉, 바운딩 박스, 물체 클래스)을 가지고 있다. 또한, 테스트 단계에서 성능이 저하될 수 있는 합성 데이터셋과 대조적으로, VOMTC는 실제 세계에서 직접 얻은 이미지 샘플을 제공하여 시간이 지남에 따른 산란광 및 주변광의 변화 등 무선 객체의 실제 시각적 특성을 포착한다.

무선 통신 연구자가 자신의 필요에 맞는 샘플을 선택하는 데 도움을 주기 위해, 데이터셋 선택 코드를 설계하였다. 이 코드를 실행할 때 사용자는 세 가지 매개변수를 조절할 수 있다.

- **활성 클래스 (active classes):** activeClasses로 나타나는 이 변수를 사용하여 각 샘플에 나타나는 클래스를 선택할 수 있다. 예를 들어, activeClasses = [2]로 설정하면 노트북만으로 구성된 RGB 및 심도 이미지를 수집할 수 있다.
- **최대 인원 수 (maximum number of people):** maxnumPeople으로 나타나는 이 매개변수를 사용하여 이미지에 나타나는 최대 사람 수를 지정할 수 있다. 예를 들어, maxnumPeople를 5로 설정하면 5명 이하의 사람이 포함된 이미지들을 얻을

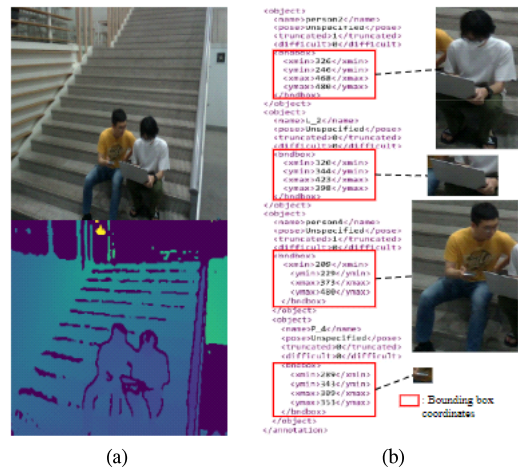


그림 3. VOMTC에서 (a) RGB 및 심도 이미지 쌍 및 (b) 해당 데이터 라벨
Fig. 3. Illustration of (a) a pair of RGB and depth images and (b) their corresponding data label in VOMTC

수 있다.

- 가장 먼 물체까지의 최대 거리 (*maximum distance to the farthest object*): \maxDist 로 나타나는 이 변수는 물체와 기지국 사이의 최대 거리를 조절한다. 예를 들어, \maxDist 를 9로 설정하여 물체까지의 거리가 9m 미만인 샘플을 수집할 수 있다.

코드를 실행하면 사용자는 원하는 S개의 RGB 및 심도 이미지의 쌍으로 구성된 데이터셋 $D \in (I^{(s)}, I_d^{(s)})_{s=1}^S$ 을 얻을 수 있다. 예를 들어, \maxnumPeople 를 각각 5 또는 1로 설정하여 다중사용자 MIMO (MU-MIMO) 또는 단일사용자 MIMO (SU-MIMO) 시스템에 적합한 샘플을 수집할 수 있다. 또한, 근거리 (near-field) 영역을 대표하는 이미지를 선택하려면 \maxDist 를 작은 값 (예: 2m)으로 설정할 수 있다. 반면, 먼 거리 (far-field) 영역을 대표하는 이미지를 원하는 경우 \maxDist 를 큰 값 (예: 37m)으로 설정할 수 있다.

IV. VOMTC를 활용한 빔포밍 기법

이 절에서는 제안하는 빔포밍 기법인 VBM의 두 가지 주요 단계 (훈련 및 테스트)에 대해 설명한다.

- VBM의 학습 단계: 완전히 새로운 물체 탐지기를 훈련시키는 대신, VOMTC를 사용하여 사전 훈련된 일반적인 물체 탐지기 (pre-trained general object detector, PT-OD)를 정밀 튜닝시킨다. 이 정밀 튜닝 단계에서는 PT-OD f_{PT-OD} 및 딥뉴럴 네트워크 (fine-tuned fully connected network, FT-FCN) f_{FT-FCN} 으로 구성된 VOMTC기반 무선 객체 탐지기가 자른 이미지 (cropped image) I_c 에서 무선 객체를 탐지하는 방법을 학습한다. 먼저, 각 물체가 특정 물체 클래스 (예: 책상, 소)에 속할 확률을 측정하는 클래스 점수를 생성한다. 그런 다음 f_{FT-FCN} 이 각 물체가 무선 객체 및 무선 객체가 아닌 확률을 측정하는 클래스 점수들인 $C_{ref,phone}(I_c)$ 와 $C_{ref,non-phone}(I_c)$ 를 생성하게끔 학습시킨다 (그림 4 참조).
- VBM의 테스트 단계: 테스트 단계에서 VBM은 VOMTC로 훈련된 무선 객체 탐지기를 사용하여 빔포밍을 수행한다. 구체적으로, 기지국은 주변 무선 환경의 이미지를 캡처하는데, 이때 이 전체 이미지 (whole image)는 두 단계의 물체 탐색을

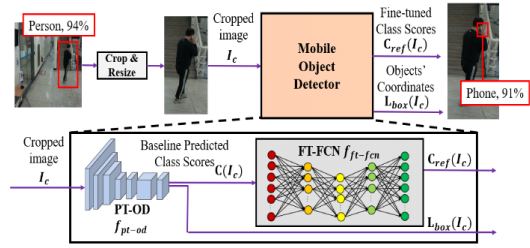


그림 4. VOMTC기반의 무선 객체 탐지기의 구조
Fig. 4. Structure of the VOMTC-based object detector

거친다. 첫 번째 단계에서는 전체 이미지에서 무선 객체를 들고 있는 사람을 먼저 탐지하기 위해 PT-OD를 사용한다. 그런 다음 두 번째 단계에서는 훈련된 무선 객체 탐지기를 사용하여 사람을 포함하는 자른 이미지에서 무선 객체를 탐지한다. VBM은 RGB 이미지에서 식별된 무선 객체의 2D 위치와 캡처된 심도 이미지를 상보적으로 활용하여 무선 객체의 방위각과 고도각 (θ, ϕ)을 얻고, 해당 방향으로 빔을 전송한다 (2.2 절을 참조).

4.1 PT-OD 및 FT-FCN을 통한 무선 객체 식별

무선 객체 탐지기의 중요한 구성 요소 중 하나는 PT-OD인데, 이는 자른 이미지에서 다양한 물체를 식별하기 위해 클래스 점수에 의존한다. 구체적으로, f_{PT-OD} 의 클래스 예측 네트워크 (class prediction network)는 각 검출된 물체가 특정 클래스에 속할 확률을 측정하는 클래스 점수 행렬 $C(I_c)$ 을 반환한다.

FT-FCN은 $C(I_c)$ 에서 선택된 정보를 기반으로 각 물체가 무선 객체에 속할 확률을 생성하는 데 사용된다. 먼저 $C(I_c)$ 에서 무선 객체와 사람이 속하는 클래스 점수 벡터들인 $C_{ref,phone}(I_c)$ 과 $C_{ref,non-phone}(I_c)$ 를 선택하고, 이 벡터들을 연결하여 입력 벡터를 생성한다. 그 후, f_{FT-FCN} 의 출력을 행렬로 변형하여 클래스 점수 행렬 $C_{ref}(I_c) = [C_{ref,phone}(I_c) \ C_{ref,non-phone}(I_c)]$ 을 얻는다.

4.2 무선 객체 탐지기의 라벨 디자인 과정

무선 객체 탐지기를 훈련시키기 위해서는 FT-FCN의 훈련 입력으로 사용될 자른 이미지들에 대한 새로운 훈련 라벨 (자른 이미지에서의 무선 객체의 바운딩 박스 좌표)이 필요하다. 이를 위해 훈련 라벨 디자인 코드 (training label design code)를 개발했으며, 이 코드는 다음과 같은 세 단계로 구성된다:

- 먼저, 데이터셋 D 에서 주어진 RGB 이미지 I^0 로부터 무선 객체를 들고 있는 사람들의 Z 개의 바운딩 박스를 식별한다. 이러한 박스들을 자르고 크

기를 조정하여 Z 개의 다른 이미지들 $\{I_c^{(s,z)}\}_{z=1}^Z$ 을 얻는다.

2. 간단한 좌표 변환을 통해 원래 이미지 $f^{(s)}$ 에서의 무선 객체 바운딩 박스 좌표 $(x_{min}, y_{min}, x_{max}, y_{max})$ 로부터 각 다른 이미지 $I_c^{(s,z)}$ 에서의 무선 객체의 바운딩 박스 좌표 $(x'_{min}, y'_{min}, x'_{max}, y'_{max})$ 를 얻는다.
3. 다른 이미지 $I_c^{(s,z)}$ 를 PT-OD의 입력으로 넣어 PT-OD가 예측한 Q 개의 바운딩 박스들의 행렬 $\mathbf{L}_{box}(I_c)$ 을 얻고 이를 $(x'_{min}, y'_{min}, x'_{max}, y'_{max})$ 와 함께 사용하여 훈련 손실을 계산한다 (4.3 절을 참조).

4.3 무선 객체 탐지기의 학습 과정

무선 객체 탐지기가 무선 객체를 정확하게 식별하기 위해, 이진 교차 엔트로피 손실 (binary cross entropy loss) L_{BCE} 을 사용하여 f_{FT-FCN} 의 네트워크 학습변수를 최적화한다. f_{FT-FCN} 의 최종 출력은 두 개의 Q -차원 벡터로 구성된 행렬이며, 이 벡터들의 q 번째 요소는 각각 q 번째 물체가 무선 객체거나 무선 객체가 아닐 확률을 나타낸다. L_{BCE} 를 계산하기 위해 이러한 벡터들을 실제 클래스 점수 벡터들인 $c_{gt\ phone}(I_c)$ 및 $c_{gt\ non-phone}(I_c)$ 와 비교해야 한다.

실제 (ground-truth) 클래스 점수 행렬 $C_{gt}(I_c) = [c_{gt,phone}(I_c)\ c_{gt,non-phone}(I_c)]$ 을 얻기 위해, 다른 이미지에서의 무선 객체의 바운딩 박스 $(x'_{min}, y'_{min}, x'_{max}, y'_{max})$ 와 $\mathbf{L}_{box}(I_c)$ 의 각 예측된 바운딩 박스 간의 Intersection-over-Union (IoU)를 계산한다. 만약 q 번째 물체의 IoU가 사전에 정의된 IoU 임계값 τ_{IoU} 를 초과한다면, 해당 물체에 클래스 'phone'을 할당하고 $c_{gt\ phone}(I_c) = 1$ 와 $c_{gt\ non\ phone}(I_c) = 0$ 으로 설정한다. 그렇지 않으면, 클래스 'non-phone'을 할당하고 $c_{gt,phone}(I_c) = 0$ 와 $c_{gt,non-phone}(I_c) = 1$ 으로 설정한다. 이후에는 이진 교차 엔트로피 손실 L_{BCE} 를 다음과 같이 구성한다.

$$L_{BCE} = - \sum_{j=1}^2 \sum_{a=1}^Q (c_{at}^{(q)} \log c_{ref\ i}^{(q)} + (1 - c_{gt\ j}^{(q)}) \log(1 - c_{ref\ j}^{(q)})) \quad (5)$$

이때, $c_{gt\ j}^{(q)}$ 와 $c_{ref\ j}^{(q)}$ 는 각각 q 번째 물체가 j 번째 클래스 ($j = 1$ 이면 무선 객체, $j = 2$ 이면 무선 객체가 아님)일 실제 확률을 예측된 확률을 나타낸다.

V. VOMTC를 활용한 빔포밍 실험

5.1 데이터셋 및 실험 설정

본 절에서는 제안하는 VBM기법의 무선 객체 탐지 및 빔포밍 성능을 모의실험을 통해 기존 기법들과 비교한다. 먼저, VBM의 f_{PD-OD} 로 기존 물체 탐지기인 EfficientDet-D8을 채택하였다. 구체적으로, 80개의 물체 클래스들과 118,000장의 학습 이미지들로 구성된 MS-COCO 2017 데이터셋^[13]을 사용하여 사전 훈련된 모델을 채택하였다.

FT-FCN은 3개의 은닉층, 입력층 및 출력층을 갖는 DNN을 사용하였다. 이 연구에서는 주어진 RGB이미지에서 상위 200개의 클래스 점수를 가진 바운딩 박스만을 고려하기 위해 식별된 바운딩 박스의 수 Q 를 200으로 설정하였다.

FT-FCN을 훈련하기 위해, activeClasses = [0,1], maxnumPeople = 6 및 maxDist = 30과 같이 세 가지 매개변수를 지정한 후 데이터셋 선택 코드를 실행하였다. 그런 다음, 이전에 선택된 7,654개의 전체 이미지들에서 (즉, $D \in (I^{(s)}, I_d^{(s)})_{s=1}^{7,654}$) 훈련 라벨 디자인 코드를 통해 9,302개의 다른 이미지들과 $\{I_d^{(s)}\}_{s=1}^{9,302}$ 그 라벨을 얻었다. VBM의 학습 단계에서는 배치 크기와 학습률을 각각 $B = 64$ 와 $\eta = 10^{-4}$ 로 설정한 후 최적화 도구인 AdamW optimizer^[14]를 사용하였다. VBM의 테스트 단계에서는 먼저 EfficientDet-D8을 사용하여 사람을 캡처한 후에 훈련된 FT-FCN을 사용하여 무선 객체를 식별하였다.

무선 객체 탐지 성능을 평가하기 위해, VOMTC 테스트 세트를 사용하였다. 성능 측정을 위해 기법의 정밀도 (precision)와 재현율(recall)을 분석하였다. 성능 비교를 위해 대규모 및 소규모 물체 탐지 단계 모두에서 EfficientDet-D8을 활용하는 CVBM^[6]과 오버샘플링 비율이 4인 8비트 Discrete Fourier Transform (DFT) 기반 빔 코드북을 사용하는 5G-BM^[15]을 선택하였다.

빔포밍 시뮬레이션에서는 $M = 64$ 개의 송신 안테나를 가진 기지국이 단일 안테나를 갖춘 무선 객체에 서비스를 제공하였다. 무선 객체는 $20 \times 20m^2$ 의 정사각형 영역에 무작위로 분포하는 다운링크 THz 시스템을 고려하였다.

경로 손실에 대해서는 통신 표준 문서인 3GPP TR 38.901 Rel.17에서 명시된 실내 경로 손실을 사용하였다. 또한, 구체적인 시스템 매개변수로는 대역폭, 반송주파수, 송신 전력 및 잡음 분산을 각각 100MHz, 0.1THz, 2W, 0.1로 설정하였다.

5.2 실험 결과

표 1에서는 VBM과 기존 CVBM의 무선 객체 탐지 성능을 제시한다. 정밀도와 재현율 측면에서 VBM이 CVBM보다 뛰어난 성능을 보이는 것을 확인할 수 있다. 예를 들어, VBM은 VOMTC 테스트 세트에서 6%의 재현율 향상을 달성하는 것을 확인할 수 있다.

물체 탐지 성과와 위치 추정 오차를 기반으로, VBM, CVBM, 5G-BM의 데이터 전송률을 사용자의 위치의 함수로 평가한다. 그림 5, 6, 7은 무선 객체의 실현 가능한 데이터 전송률을 나타내는 3D 색상 지도이다. 실험 결과, VBM이 대부분의 서비스 영역에서 기존 빔 관리 기법들을 능가함을 보여주며, 5G-BM과 CVBM 대비 각각 36%와 14%의 데이터 전송률 향상을 달성한다.

제안하는 VBM 기법이 기존 빔포밍 기법들보다 높은 데이터 전송률을 제공하지만, 캡처된 이미지에서 객체 탐지가 제대로 작동하지 않을 수 있다. 예를 들어, 장애물에 의해 무선 객체가 시각적으로 차단될 때, VBM을 사용하는 기지국은 빔 방향을 정확히 식별하지 못할 수 있다.

이러한 문제를 해결하기 위해 세 가지 주요 접근 방

표 1. VOMTC 테스트 세트에서 평가한 무선 객체 탐지 및 위치 추정 성능
Table 1. Phone detection and localization performance evaluated on the VOMTC Test set

	Cell Phone		Localization Error	
	Precision	Recall	Azimuth Angle (rad)	Elevation Angle (rad)
VBM	94.46	81.04	0.0804	0.0804
CVBM	94.17	76.78	0.0972	0.0973

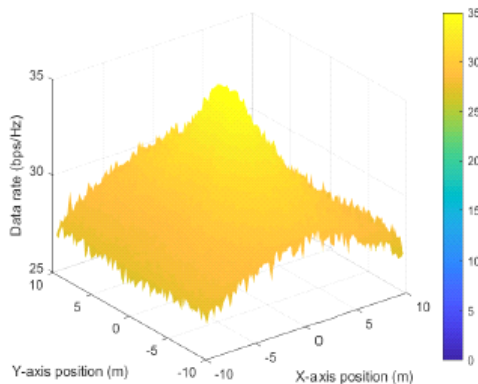


그림 5. (0,0)에 위치한 높이 3m인 기지국이 VBM을 썼을 때 각 사용자 위치 (x,y)에서의 데이터 전송률
Fig. 5. Illustration of data rate at each user position (x,y) when the 3m-high BS located at (0,0) uses VBM

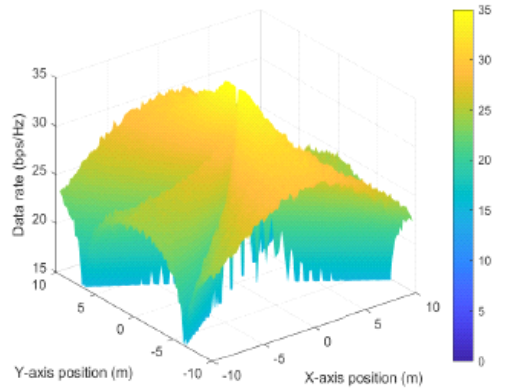


그림 6. (0,0)에 위치한 높이 3m인 기지국이 CVBM을 썼을 때 각 사용자 위치 (x,y)에서의 데이터 전송률
Fig. 6. Illustration of data rate at each user position (x,y) when the 3m-high BS located at (0,0) uses CVBM

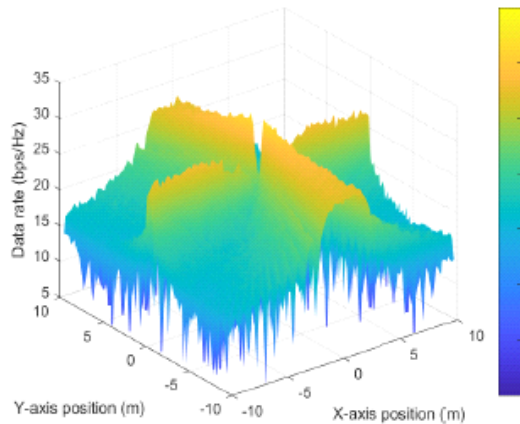


그림 7. (0,0)에 위치한 높이 3m인 기지국이 5G-BM을 썼을 때 각 사용자 위치 (x,y)에서의 데이터 전송률
Fig. 7. Illustration of data rate at each user position (x,y) when the 3m-high BS located at (0,0) uses 5G-BM

법을 고려할 수 있다. 첫째, VBM과 기존 무선 신호 전송 방식을 혼합하여 사용하는 방법이다. 구체적으로, 무선 객체가 이미지에서 보이는 상황에서는 VBM을 사용하고, 장애물이 VBM을 통해 감지되면 기존의 빔 인덱스 피드백 및 코드북 기반 빔포밍 방식으로 전환할 수 있다. 둘째, 여러 센서 장치(예: 레이더, 초음파 센서, 적외선 및 열화상 카메라)로부터 얻은 다중 모달 센싱 정보를 활용하는 방법이다. 예를 들어, 레이더와 카메라를 동시에 사용하여 장애물에 의해 차단된 대상 무선 객체를 탐지할 수 있다. 마지막으로, 다양한 각도에서 촬영된 여러 장의 이미지를 종합적으로 활용하여 무선 객체의 위치 추정 성능을 향상시킬 수 있다. 예를 들어, 한쪽 각도에서 촬영한 이미지에서 보이지 않던 무선 객

체가 다른 각도에서 촬영한 이미지에서는 나타날 수 있다. 이를 통해 VBM을 사용하여 무선 객체의 위치를 정확히 파악하고, 궁극적으로 빔 방향을 정밀하게 식별할 수 있다. 이러한 접근 방법들은 VBM의 단점을 미래에 개선하는 데 기여할 수 있다.

또한, 무선 객체가 빠르게 이동하는 경우 채널 코히런스 타임이 CSI-RS 파일럿 신호 송신 주기보다 짧아질 수 있다. 이로 인해 파일럿 신호가 송신되지 않는 시간 동안 데이터를 전송하면 실제 채널과 추정된 채널 간에 큰 차이가 발생하는 채널 에이징 효과(channel aging effect)가 나타날 수 있다. 이 경우, 기지국이 무선 객체의 실제 방향으로 정확한 빔포밍을 수행하지 못해 데이터 전송 효율이 크게 저하될 수 있다.

이 문제를 해결하기 위해, 캡처한 이미지로부터 무선 객체 탐지기를 사용하여 무선 객체의 각도와 거리를 빠르게 파악한 후, 수신된 파일럿 신호를 기반으로 선형적 알고리즘(예: 최소 자승법)을 사용하여 채널의 페이딩 계수만을 추정한다. 이렇게 획득한 채널 파라미터들인 각도, 거리 및 페이딩 계수를 이용하여 전체 채널을 효과적으로 복원할 수 있으며, 정확한 채널 정보를 기반으로 데이터 전송 시 빔포밍을 최적화할 수 있다. 따라서 향후 연구에서는 CSI 무선 신호와 이미지 센싱 정보를 공동으로 활용하여 빔포밍의 정확성과 효율성을 향상시키는 방법에 대한 논의가 필요할 것이다.

VI. 결 론

본 논문에서는 컴퓨터 비전 기반 빔포밍을 수행하기 위해 특화된 대규모 비전 데이터셋을 소개하였다. 소개한 VOMTC의 효과를 입증하기 위해 VBM이라는 컴퓨터 비전 기반 빔포밍 기술을 개발하였다. 무선 객체 탐지 및 빔포밍에 대한 실험 결과로부터, VBM이 기존의 빔포밍 기술보다 우수함을 확인할 수 있었다. 본 연구에서 제안한 기법은 향후 빔 트래킹 (beam tracking) 및 무선 객체 위치 측정 (positioning)와 같은 다양한 무선 통신 분야에 사용될 것으로 기대한다.

References

[1] Y. Ahn, J. Kim, S. Kim, S. Kim, and B. Shim, "Sensing and computer vision-aided mobility management for 6G millimeter and terahertz communication systems," *IEEE Trans. Commun.*, 2024. (<https://doi.org/10.1109/TCOMM.2024.339279>)

9)

- [2] Q. Xue, C. Ji, S. Ma, J. Guo, Y. Xu, Q. Chen, and W. Zhang, "A survey of beam management for mmWave and THz communications towards 6G," *IEEE Commun. Surv. Tut.*, 2024. (<https://doi.org/10.1109/COMST.2024.3361991>)
- [3] S. Jeong, H. Ju, S. Kim, and B. Shim, "Automated environment-aware channel feedback for 6G massive MIMO systems," in *Proc. IEEE Glob. Commun. Wkshps.*, 2023. (<https://doi.org/10.1109/GCWkshps58843.2023.10465185>)
- [4] B. Ning, Z. Tian, W. Mei, Z. Chen, C. Han, S. Li, J. Yuan, and R. Zhang, "Beamforming technologies for ultra-massive MIMO in terahertz communications," *IEEE Open J. Commun. Soc.*, vol. 4, pp. 614-658, Feb. 2023. (<https://doi.org/10.1109/OJCOMS.2023.3245669>)
- [5] S. Kim and B. Shim, "Frequency-dependent precoding based beam training in terahertz massive MIMO systems," *J. KICS*, vol. 47, no. 10, pp. 1535-1537, 2022. (<https://doi.org/10.7840/kics.2022.47.10.1535>)
- [6] C. Han, L. Yan, and J. Yuan, "Hybrid beamforming for terahertz wireless communications: Challenges, architectures, and open problems," *IEEE Wireless Commun.*, vol. 28, no. 4, pp. 198-204, Aug. 2021. (<https://doi.org/10.1109/MWC.001.2000458>)
- [7] Y. Ahn, J. Kim, S. Kim, K. Shim, J. Kim, S. Kim, and B. Shim, "Towards intelligent millimeter and terahertz communication for 6G: Computer vision-aided beamforming," *IEEE Wireless Commun.*, vol. 30, no. 5, pp. 179-186, Oct. 2023. (<https://doi.org/10.1109/MWC.007.2200155>)
- [8] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," in *arXiv preprint arXiv:1804.02767*, 2018. (<https://doi.org/10.48550/arXiv.1804.02767>)
- [9] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in

Proc. CVPR, pp. 10781-10790, 2020.

(<https://doi.org/10.1109/CVPR42600.2020.01079>)

- [10] S. Kim, J. Park, J. Moon, and B. Shim, "Fast and accurate terahertz beam management via frequency-dependent beamforming," *IEEE Trans. Wireless Commun.*, vol. 23, no. 3, pp. 1699-1712, Mar. 2024.
(<https://doi.org/10.1109/TWC.2023.3291440>)
- [11] W. Kim, Y. Ahn, and B. Shim, "Deep neural network-based active user detection for grant-free NOMA systems," *IEEE Trans. Commun.*, vol. 68, no. 4, pp. 2143-2155, Apr. 2020.
(<https://doi.org/10.1109/TCOMM.2020.2969184>)
- [12] S. Kim, J. Moon, J. Kim, Y. Ahn, D. Kim, S. Kim, K. Shim, and B. Shim, "Role of sensing and computer vision in 6G wireless communications," *IEEE Wireless Commun.*, 2024.
(<https://doi.org/10.1109/MWC.016.2300526>)
- [13] T. Y. Lin, et al., "Microsoft COCO: Common objects in context," in *Proc. ECCV*, 2014.
(https://doi.org/10.1007/978-3-319-10602-1_48)
- [14] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017.
(<https://doi.org/10.48550/arXiv.1711.05101>)
- [15] Y. Li, B. Gao, X. Zhang, and K. Huang, "Beam management in millimeter-wave communications for 5G and beyond," *IEEE Access*, vol. 8, pp. 13282-13293, Jan. 2020.
(<https://doi.org/10.1109/ACCESS.2019.2963514>)

김 선 우 (Sunwoo Kim)



2018년 : Stony Brook University Electrical Engineering 학사 졸업

2018년~현재 : 서울대학교 전기정보공학과 박사과정

<관심분야> 딥러닝, 무선통신, 신호처리

[ORCID:0000-0003-2622-4136]

김 재 우 (Jaewoo Kim)



1996년 : 한양대학교 전자통신공학과 학사 졸업

2010년 : 한양대학교 전자컴퓨터통신공학과 석사 졸업

1996년~2005년 : 삼성전자 SOC 개발실

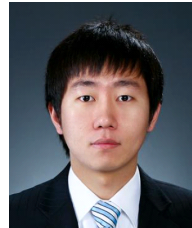
2005년~2010년 : 포스코ICT Wibro 연구실

2010년~현재 : 한화시스템 수석연구원

<관심분야> 5G/6G 이동통신, 전술통신, 기계학습 기반 통신 시스템

[ORCID:0000-0002-2012-812X]

이 종 만 (Jongman Lee)



2010년 : 경희대학교 전기무선공학과 학사 졸업

2022년 : 연세대학교 전기전자공학과 석사 졸업

2010년~현재 : 한화시스템 전문연구원

<관심분야> 전술통신시스템, 미래전술통합단말기, 전술 IoT, AI 기반 SON(Self Organizing Network)

[ORCID:0000-0002-8467-6504]

심 병 효 (Byonghyo Shim)



1995년: 서울대학교 제어계측
공학과 학사 졸업

1997년: 서울대학교 제어계측
공학과 석사 졸업

2004년: University of Illinois
at Urbana-Champaign Ma-
thematics 석사 졸업

2005년: University of Illinois
at Urbana-Champaign Electrical and Computer
Engineering 박사 졸업

2005년~2007년: 미 쉘컴사 선임 및 책임연구원

2007년~2014년: 고려대학교 컴퓨터 전파통신공학과
부교수

2014년~현재: 서울대학교 전기정보공학과 정교수

<관심분야> 정보이론, 무선통신, 신호처리

[ORCID:0000-0001-5051-1763]