

2023 전기 졸업과제 최종보고서

물류창고에서 이동체 위치 추적을 위한
BLE Zoning 시스템 개발



201824409 강중헌
201824454 김지명
201824556 이정현

지도교수 정상화

목 차

1. 과제 개요	1
1.1 과제 배경	1
1.2 과제 목표	1
1.3 과제 내용 및 요구사항	2
2. 과제 구성 및 배경지식	2
2.1 시스템 구성도	2
2.2 모델 구성	3
2.3 배경지식	6
2.4 실험환경	13
2.5 개발환경	14
3. 과제 설계 상세	15
3.1 데이터 셋	15
3.2 모델구성	17
3.3 웹서비스 구현	19
4. 과제 수행 내용 및 결과	24
4.1 Data filtering	24
4.2 커널 선택	25
4.3 데이터 증강(data augmentation)	26
4.4 Zoning System	29
5. 결론 및 향후 연구 방향	34
5.1 결론	34
5.2 산학협력 멘토링	35

5.3 향후 연구 방향	35
6. 개발 일정 및 역할 분담	36
6.1 개발 일정	36
6.2 역할 분담	36
7. 참고 문헌	37

1. 과제 개요

1.1 과제 배경

물류창고에서 자재 이동 및 근로자의 움직임을 효율적으로 관리하고 추적하는 것은 운영 최적화와 생산성 향상에 중요하다. 물류 및 창고 관리 분야에서 실시간 위치 추정 시스템은 물품 및 근로자의 정확한 추적을 가능케 하여 생산성과 효율성을 향상시키는 데 중요한 역할을 한다. 기존의 위치 추정 기술들은 정확도와 비용, 설치의 어려움 등 여러 가지 한계점을 지니고 있다. 기존 방식들의 한계점을 극복하기 위해, 최근에는 BLE(Bluetooth Low Energy) 기술을 활용한 실시간 위치 추정 시스템을 향한 관심이 증가하고 있다.

물류 및 창고 관리 분야에서 BLE(Bluetooth Low Energy) 기술은 저 에너지 소모, 저렴한 비용, 정확한 내부 위치 파악 능력 등 다양한 장점을 제공한다. 이러한 장점들을 중심으로, 물류 작업의 정확한 추적과 근로자의 움직임을 실시간으로 관리하는 필요성이 대두되고 있다.

1.2 과제 목표

본 과제는 위 방식들의 한계점을 극복하기 위해 BLE, 지도학습, 핑거프린트(Fingerprint) 방식 등을 활용하여 이동체 추정 Zoning 시스템 구현을 목표로 한다.

1.3 과제 내용 및 요구사항

BLE Zoning 기술을 활용하여, 물류창고에서 자재와 근로자의 실시간 이동 추정 기술을 구현하고 분석.

< 네트워크 메트릭을 활용하여 BLE 노드의 위치를 추정하는 학습기법 개발>

-수집한 RSSI(Radio Signal Strength Indicator) 값을 이용하여 지도학습 모델을 구성하고, 이를 활용하여 물체의 위치를 추정.

< 물류 환경의 실시간 이동체 관리시스템 개발 및 추가적 서비스 도출>

- 실시간으로 웹에서 이동하는 위치를 표시하는 맵을 나타내 BLE zoning 시스템을 구현.

2. 과제 구성 및 배경지식

2.1 시스템 구성도

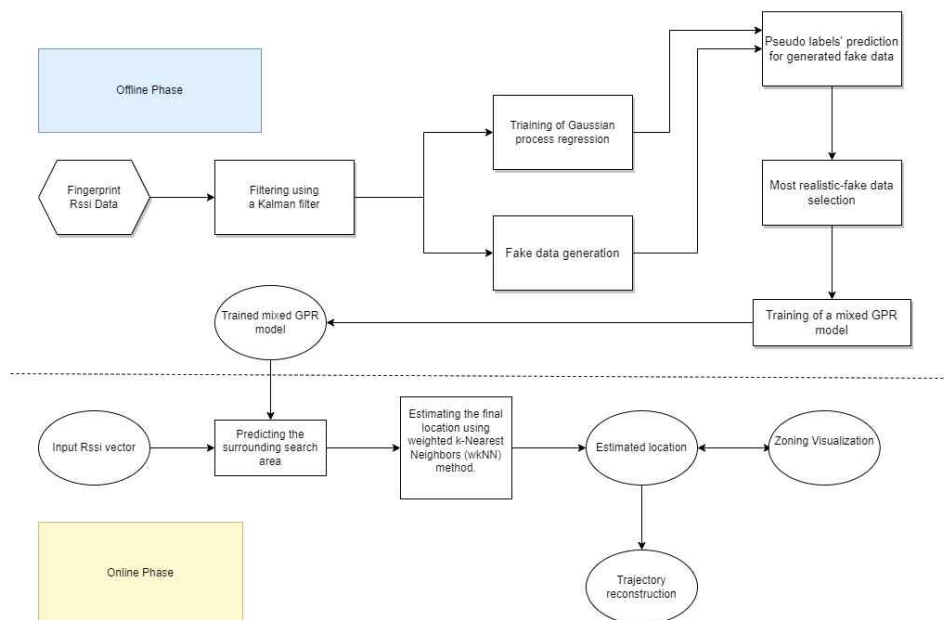


그림 1 전체 시스템 flow

2.2 모델 구성

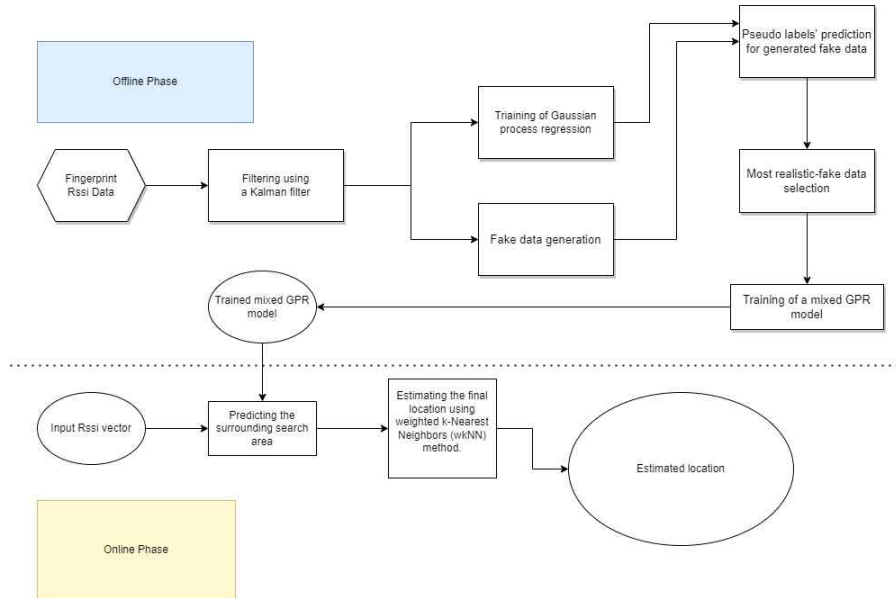


그림 2 model

지도학습 모델 생성 시 오프라인 단계 (Offline Phase), 온라인 단계(Online Phase) 두 가지 단계로 구분한다.

<오프라인 단계-Offline Phase>

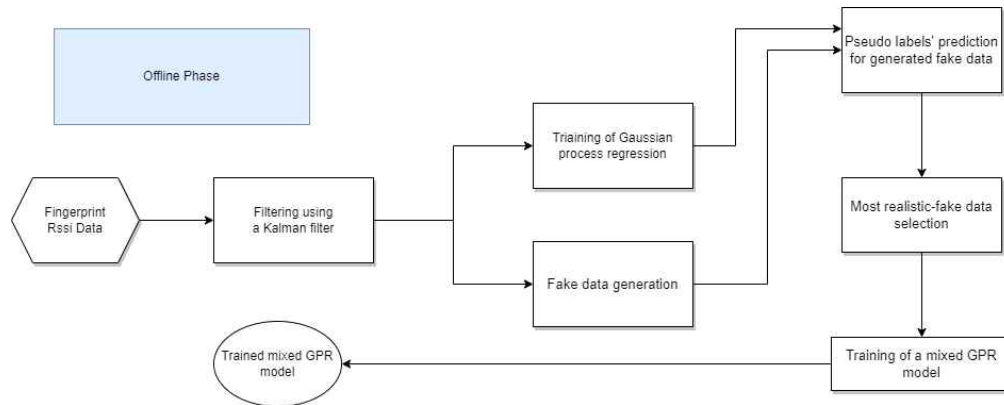


그림 3 offline phase

오프라인 단계에서는 훈련 및 초기화 단계로, 데이터의 수집과 모델의 학습, 최적화가 이루어지는 단계를 나타낸다. 이 단계에서는 모델이 대량의 훈련 데이터를 사용하여 학습하고, 최적의 모델 파라미터를 찾아낸다. 또한, 학습된 모델의 성능을 평가하고 모델을 조율한다. 이후 학습된 모델이 온라인 단계에서 사용될 준비가 된다.

본 과제에서 오프라인 단계는 RSSI 벡터 지문데이터(labeled RSSI vector fingerprint data)를 수집한 후 RSSI 벡터 지문 데이터의 풍부함을 강화하기 위해 GAN을 사용하여 fake RSSI 벡터 지문 데이터를 생성하고 지도학습 모델을 학습, 훈련하고 최적의 모델을 찾는다.

데이터 수집 및 전처리 단계에서 레이블이 지정된 실내 환경의 RSSI 지문 데이터를 수집하고, GPR(Gaussian Process regression)을 활용하여 데이터를 확장하고 다양성을 보장하는 fake data를 생성한다. 이후, 지도학습(GPR)을 통해 의사라벨을 할당하고 불필요하거나 부정확한 데이터를 정제하여 최종적으로 정확하고 다양한 데이터 세트를 얻게 된다. 온라인 단계에서는 이러한 최적화된 모델을 사용하여 동적 환경에서 발생하는 데이터를 기반으로 정확한 위치 추정을 제공한다.

<온라인 단계-Online Phase>

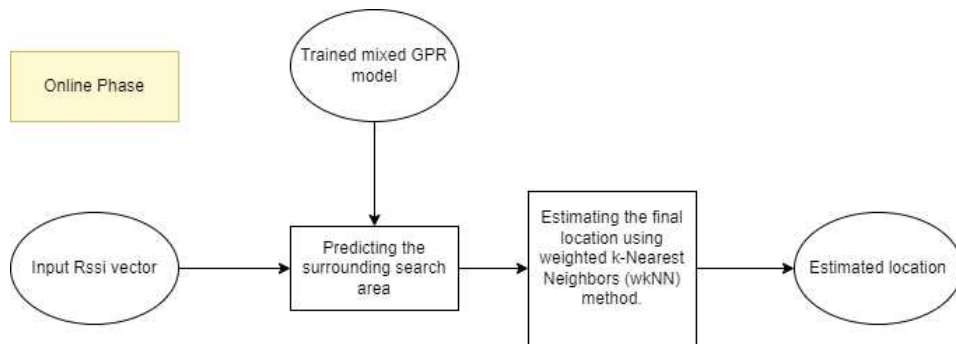


그림 4 online phase

온라인 단계(Online Phase)에서는 미리 학습된 GPR 모델을 활용하여 실시간으로 수집된 RSSI 벡터값을 기반으로 위치를 추정하는 단계이다. 이 단계에서 모델은 새로운 데이터를 받으면, 즉시 위치를 추정하고, 그 결과를 사용자에게 반환하거나 다른 응용 프로그램에 전달한다.

본 과제에서는 지도학습 GPR 와 wkNN(Weighted k-Nearest Neighbors algorithm)을 결합하여 위치를 추정한다. 먼저 GPR 모델을 사용하여 주변 검색 영역을 예측하고, 그 후에 wkNN을 사용하여 예측된 영역 내에서 k 개의 최근접 이웃을 찾는다. 각 이웃과 추정 위치 간의 거리에 대한 가중치를 계산하고, 이 가중치를 사용하여 이웃들의 값에 가중평균을 적용하여 최종 위치를 추정한다. 이러한 방식으로 GPR의 불확실성 처리 능력과 wkNN의 정확도를 결합하여 실시간 위치 추정을 수행할 수 있다.

2.3 배경지식

2.3.1 BLE

Bluetooth Low Energy(BLE) 무선 통신은 매우 낮은 전력으로 작동하도록 설계 되었다. 처음에는 장치 통신 기능으로 알려졌던 BLE 는 이제 고정밀 실내 위치 서비스에 대한 수요 증가를 해결하기 위한 장치 위치 확인 기술로도 널리 사용되고 있다. 이제 BLE 에는 한 장치에서 다른 장치의 존재 여부, 거리(RSSI), 방향을 확인할 수 있는 기능이 포함되어 있다.¹

	Bluetooth Low Energy (LE)	Bluetooth Classic
Frequency Band	2.4GHz ISM Band (2.402 – 2.480 GHz Utilized)	2.4GHz ISM Band (2.402 – 2.480 GHz Utilized)
Channels	40 channels with 2 MHz spacing (3 advertising channels/37 data channels)	79 channels with 1 MHz spacing
Channel Usage	Frequency-Hopping Spread Spectrum (FHSS)	Frequency-Hopping Spread Spectrum (FHSS)
Modulation	GFSK	GFSK, $\pi/4$ DQPSK, 8DPSK
Data Rate	LE 2M PHY: 2 Mb/s LE 1M PHY: 1 Mb/s LE Coded PHY (S=2): 500 Kb/s LE Coded PHY (S=8): 125 Kb/s	EDR PHY (8DPSK): 3 Mb/s EDR PHY ($\pi/4$ DQPSK): 2 Mb/s BR PHY (GFSK): 1 Mb/s
Tx Power*	≤ 100 mW (+20 dBm)	≤ 100 mW (+20 dBm)
Rx Sensitivity	LE 2M PHY: ≤ -70 dBm LE 1M PHY: ≤ -70 dBm LE Coded PHY (S=2): ≤ -75 dBm LE Coded PHY (S=8): ≤ -82 dBm	≤ -70 dBm
Data Transports	Asynchronous Connection-oriented Isochronous Connection-oriented Asynchronous Connectionless Synchronous Connectionless Isochronous Connectionless	Asynchronous Connection-oriented Synchronous Connection-oriented
Communication Topologies	Point-to-Point (including piconet) Broadcast Mesh	Point-to-Point (including piconet)
Positioning Features	Presence: Advertising Direction: Direction Finding (AoA/AoD) Distance: RSSI, HADM (Coming)	None

그림 5 Bluetooth Low Energy vs Bluetooth Classic

¹ Bluetooth Technology Overview . (2023, Oct 13).
<https://www.bluetooth.com/learn-about-bluetooth/tech-overview/>

2.3.2 RSSI

무선통신에서 RSSI(radio signal strength indicator)는 전송된 무선 신호의 강도를 나타내며, 주변 환경, 거리, 장비 특성에 영향을 받는다. 높은 RSSI 값은 강한 신호를 나타내며, 낮은 값은 약한 신호를 나타낸다. 이러한 RSSI 는 무선 통신 시스템의 성능을 평가하고 효율적인 통신을 위한 중요한 지표로 활용된다.

2.3.3 Fingerprinting

핑거프린트(Fingerprinting) 방식은 무선 신호 강도(RSSI) 값을 기반으로 실내 위치를 추정하는 기술이다. 이 방식은 주변 무선 액세스 포인트(APs) 또는 블루투스 Beacon 과 같은 무선 장치들의 신호 강도를 미리 측정하고 데이터베이스에 저장해 둔다.

이러한 미리 측정된 데이터는 위치를 추정할 때 사용된다. 추정하려는 위치에서 주변의 무선 신호 강도를 측정하고, 이 값을 미리 저장된 데이터베이스에 있는 핑거프린트와 비교하여 가장 유사한 핑거프린트를 찾는다. 그런 다음 해당 핑거프린트에 연결된 위치 정보를 사용하여 실내 위치를 추정할 수 있다.

핑거프린트 방식은 주로 실내 위치 기반 서비스, 실내 내비게이션, 스마트 홈 시스템 등의 응용 프로그램에서 사용된다. 이 방식은 무선 신호의 강도 변화에 영향을 받을 수 있지만, 정확한 데이터베이스 구축과 적절한 알고리즘을 사용하면 정확한 위치 추정이 가능하다.

2.3.4 GPR

Gaussian Process Regression (GPR)은 확률론적 회귀 모델로, 입력 공간의 모든 점에서 연속적인 출력을 예측하는 비모수적(non-parametric) 방법이다. GPR 은 주어진 데이터 포인트의 불확실성을 포함한 모든 가능한 함수에 대한 확률 분포를 추정한다. 이는 기존의 선형 회귀 모델과 달리, 데이터에 대한 가정이나 함수 형태에 대한 제약이 없다는 장점을 가지고 있다. GPR은 데이터 간의 상관 관계를 포착하고, 예측의 불확실성을 제공함으로써 특히 잡음(noise)이 많은 데이터나 적은 데이터로부터 효과적으로 학습할 수 있다.

GPR의 핵심 아이디어는 입력 공간의 각 점을 Gaussian 분포로 표현하고, 주어진 데이터로부터 이러한 분포의 파라미터를 조절하는 것이다. 이를 통해 새로

운 입력값에 대한 예측은 주어진 데이터 포인트들과의 상호 작용을 기반으로 한 확률 분포로 계산된다. 또한, GPR은 하이퍼파라미터 조정이 필요한데, 이는 모델의 성능에 큰 영향을 미친다. 따라서 적절한 하이퍼파라미터 조정을 통해 GPR 모델을 최적화하는 것이 중요하다.

GPR은 SVM이나 Linear Regression에 비해 불확실성, 비선형성을 modeling 함에 강점이 있어 본 과제에서 사용하였다.

2.3.5 wkNN

Weighted k-Nearest Neighbors (wkNN)는 k-최근접 이웃 알고리즘(KNN)의 확장으로, 주변 이웃 데이터 포인트의 가중치를 고려하여 예측을 수행하는 비모수적인 ML 알고리즘이다. wkNN은 새로운 데이터 포인트에 대한 예측을 할 때, 주변 k개의 이웃 데이터 포인트를 찾고 이 이웃들의 가중평균을 사용하여 예측을 수행한다. 이때, 각 이웃 데이터 포인트의 가중치는 일반적으로 거리에 기반하여 계산되며, 더 가까운 이웃에 더 높은 가중치를 부여함으로써 모델의 정확성을 향상할 수 있다.

wkNN은 다양한 분야에서 패턴 인식, 회귀, 분류 등의 작업에 사용된다. 주로 거리 측정과 가중치 계산 방법의 조절을 통해 모델의 성능을 조정할 수 있어 유연성을 가지고 있다. 또한, 레이블(label)되지 않은 데이터에 대한 예측이 필요한 경우에도 효과적으로 적용될 수 있어 실전 문제에 유용한 알고리즘 중 하나이다.

2.3.6 Kalman filter

칼만 필터(Kalman Filter)는 시간에 따라 변화하는 시스템의 상태를 추정하는데 사용되는 재귀적인 수학적 알고리즘이다. 주로 시스템의 상태, 측정 값, 시스템의 동적 모델, 그리고 노이즈에 대한 정보를 활용하여 정확한 예측을 수행한다.

칼만 필터는 시스템의 현재 상태에 대한 추정치를 제공하고 새로운 측정값을 받을 때마다 이 추정치를 업데이트한다. 이 과정은 시스템의 상태에 대한 불확실성을 고려하면서 진행된다. 칼만 필터는 예측 단계와 업데이트 단계로 구성되어 있다.

<예측 단계(Prediction Step)>

- 이전 상태 추정치를 기반으로 시스템이 어떻게 변할 것인지 예측
- 시스템의 동적 모델이 사용됨

<업데이트 단계(Update Step)>

- 예측치와 실제 측정치 간의 차이를 이용하여 시스템의 현재 상태를 업데이트
- 노이즈와 불확실성을 고려하여 상태를 조정

칼만 필터는 실시간 데이터를 다루거나, 노이즈가 있는 환경에서 사용될 때 특히 유용하다. 예를 들어, 자율 주행 자동차의 위치 추정, 로봇의 자세 추정, 항법 시스템, 신호 처리 등 다양한 분야에서 활용된다.

2.3.7 Kernel Function

커널 함수는 머신 러닝 모델에서 데이터 간의 유사도를 계산하는 함수이다. 주로 서포트 벡터 머신(SVM)과 가우시안 프로세스(GP)와 같은 알고리즘에서 사용된다. 커널 함수는 데이터를 고차원 특징 공간으로 매핑하여 비선형 관계를 모델링하는 데 도움을 주는 역할을 한다.

가우시안 프로세스(Gaussian Process) - 주요 커널 함수

<Gaussian Radial Basis Function Kernel>

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right)$$

수식 1 RBF kernel

이 커널은 데이터 간의 거리에 기반하여 유사도를 계산하며, 가우시안 분포의 형태를 모방한다. 데이터를 고차원 공간으로 매핑하여 비선형 관계를 모델링할 수 있다. σ 는 가우시안 커널의 대표적인 매개변수로, 커널의 폭을 조절한다.

<Matern Kernel>

$$K(x, y) = \frac{1}{2^{v-1} \Gamma(v)} \left(\sqrt{2v} \frac{\|x - y\|}{l} \right)^v B_v \left(\sqrt{2v} \frac{\|x - y\|}{l} \right)$$

수식 2 Matern kernel

여기서 v 는 정수 또는 반정수이고, B_v 는 modified Bessel 함수이다. 이 커널은 부드러운 특성을 가지며, v 값에 따라서 선형부터 매우 비선형까지 다양한 패턴을 모델링 할 수 있다.

<Quadratic Kernel>

Quadratic Kernel은 주어진 데이터의 비선형 관계를 모델링하기 위해 사용된다.

$$K(x, y) = (x \cdot y + c)^2$$

수식 3 Quadratic Kernel

이 커널은 주어진 두 데이터 포인트의 내적값에 제곱을 적용하여 계산한다. 여기서 x 와 y 는 입력 데이터 포인트이고, c 는 상수이다. Quadratic Kernel은 주어진 데이터가 다항식 함수로 나타내어질 수 있는 '비선형 관계를 가지고 있다.'라 가정한다. 이 커널은 데이터 간의 상호 작용을 포착하고 복잡한 비선형 패턴을 모델링 하는데 사용된다.

2.3.8 베이지안 최적화(Bayesian Optimization)

베이지안 최적화(Bayesian Optimization)는 함수의 최적 값을 찾는 데 사용되는 최적화 알고리즘 중 하나이다. 주어진 함수의 출력 값을 최소화 또는 최대화 하는 최적의 입력값을 찾는 것이 목표이다. 이를 통해 시간과 자원을 효율적으로 사용하여 최적의 하이퍼파라미터 조합이나 모델 구조를 찾을 수 있다.

베이지안 최적화는 다음과 같은 주요 특징을 가지고 있다

<확률적 모델 사용>

베이지안 최적화는 주어진 함수의 값을 모델링하기 위해 확률적인 모델을 사용한다. 일반적으로, 가우시안 프로세스(Gaussian Process) 모델이 사용된다. 이 모델은 현재까지 관찰된 데이터를 기반으로 함수의 형태를 확률적으로 예측한다.

<베이지안 추론>

이전에 관찰된 데이터와 새로운 데이터의 불확실성을 고려하여 다음 번 시도 할 후보적인 입력 값을 선택한다. 이를 통해 불필요한 시도를 피하고, 더 확실한 지점으로 진행한다.

<자원 효율적>

베이지안 최적화는 몇 번의 시도로도 함수의 최적값을 찾는 데 효과적이다. 불필요한 시도를 피하고 목적 함수를 적은 횟수로 최적화할 수 있다.

<조기 종료 기능>

베이지안 최적화는 반복적으로 함수를 호출하며, 성능이 더 이상 향상되지 않으면 조기에 최적화를 중단할 수 있다.

이러한 특성들로 인해 베이지안 최적화는 머신 러닝 모델의 하이퍼파라미터 튜닝, 신경망 아키텍처 최적화, 실험 계획 등 여러 영역에서 널리 사용된다. 주로 `scikit-optimize`, `BayesianOptimization`과 같은 라이브러리를 사용하여 구현할 수 있다.

2.3.9 GAN(Generative Adversarial Network) 과 데이터 증강 (Data augmentation)

GAN(Generative Adversarial Network)은 Deep Learning의 한 분야로, 생성 모델(Generative Model)을 학습하는 데 사용되는 인공 신경망 구조이다. GAN은 두 개의 신경망, 즉 생성자(Generator)와 판별자(Discriminator)로 구성된다. 이 두 신경망은 서로 경쟁적인 과정을 통해 학습된다. GAN을 사용한 데이터 증강은 실제 데이터 분포와 유사한 fake data를 생성하여 기존 데이터셋을 확장하는 기술이다. 이 기술은 주어진 문제에 대한 데이터 양을 증가시키고, 모델의 일반화 성능을 올리기 위해 사용된다. GAN을 사용한 데이터 증강은 다음과 같은 방식으로 진행한다.

<생성자(Generator) 학습>

생성자 네트워크는 무작위 노이즈 벡터를 입력으로 받아 실제 데이터와 유사한 데이터를 생성한다.

<판별자(Discriminator) 학습>

판별자 네트워크는 실제 데이터와 생성자가 생성한 fake data를 구별하도록 학습한다. 판별자는 실제 데이터와 생성된 데이터를 정확하게 분류할 수 있도록 훈련된다.

<경쟁적 학습>

생성자와 판별자는 서로 경쟁적인 관계에 있어 생성자는 판별자를 속이도록, 판별자는 생성자에게서 온 fake data를 정확하게 구분하도록 학습된다. 이러한 경쟁적 학습은 생성자가 더 실제 데이터와 유사한 샘플을 생성하고, 판별자는 더 정확하게 판별하게끔 이끈다.

<샘플 생성과 데이터 증강>

훈련된 생성자는 무작위 노이즈를 입력으로 받아 실제 데이터와 유사한 fake data를 생성할 수 있다. 이러한 fake data는 기존의 실제 데이터와 결합하여 더 큰 규모의 데이터 셋을 형성함으로써 데이터 양을 증가시키는 데 사용된다.

이를 통해 모델은 더 많은 다양성과 일반화 능력을 갖추게 되어 새로운 데이터에 대한 예측 성능이 향상될 수 있다.

2.4 실험환경

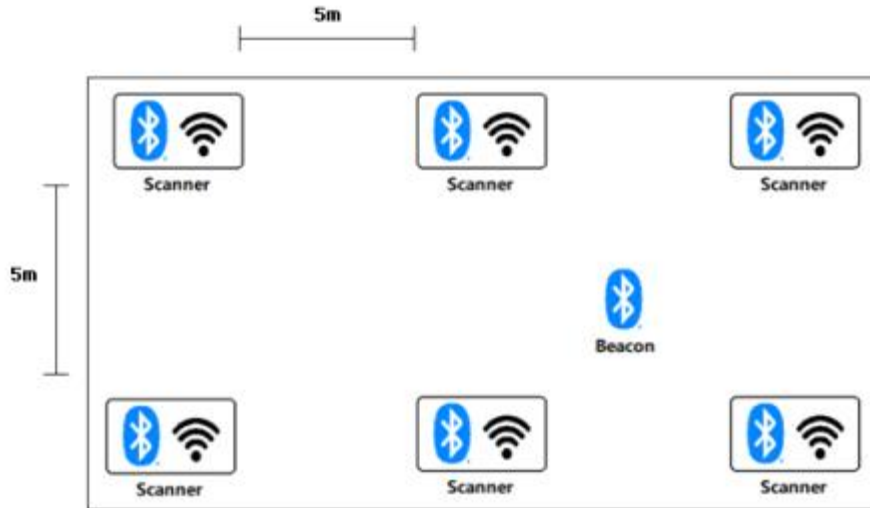


그림 6 비콘, 스캐너 위치 배치-실험환경'

실험환경은 가로 10m, 세로 5m의 실내 공간에 5m 간격으로 총 6개의 스캐너를 설치하였다. 각 스캐너는 다음 위치에 배치되었다: (0,0), (5,0), (10,0), (0,5), (5,5), (10,5). 이 스캐너들은 Beacon에서 발신되는 BLE 신호 RSSI 값을 수집하고 이 정보를 데이터베이스로 송신한다. 실험에서 사용된 스캐너와 Beacon은 모두 Lolin d32 보드를 이용하였고, 이 실험은 실내 공간에서 진행되었다.

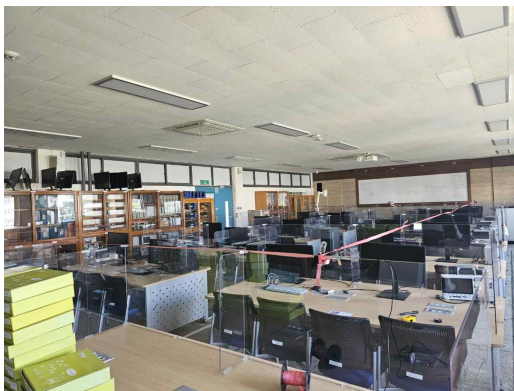


그림 7 실험환경 실내 내부 사진-1

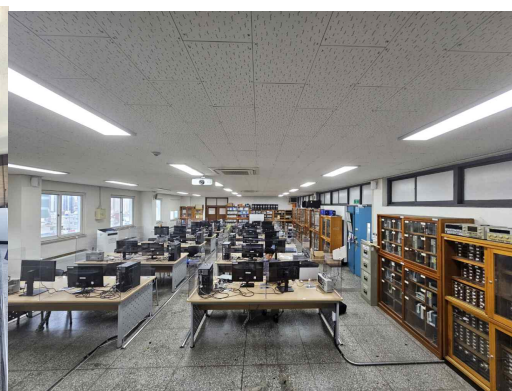


그림 8 실험환경 실내 내부 사진-1

2.5 개발환경

2.5.1 개발언어

Arduino C/C++ 코드를 Lolin d32 보드에 업로드한다. 데이터베이스와의 연동 및 ML 모델 개발에도 Python을 이용한다.

2.5.2 개발도구

Arduino IDE를 사용하여 Lolin d32 보드에 대한 코드를 작성한다. 데이터 수집 및 전처리를 위해 Python의 라이브러리를 사용하였으며, 데이터베이스와의 연동에는 HeidiSQL을 사용한다. ML 모델 개발에는 Python의 주요 라이브러리인 scikit-learn을 사용하여 Gaussian Process Regression 모델을 구현한다.

2.5.3 데이터베이스(db)

MariaDB 데이터베이스를 사용하기 위해 HeidiSQL을 통해 스키마를 설계하고 데이터를 저장한다. 데이터는 WiFi 연결을 통해 스캐너(Lolin d32 보드)에서 수집되어 MariaDB 데이터베이스에 저장된다.

2.5.4 실시간 위치 추정 및 시각화

수집된 RSSI 값들을 기반으로 실시간으로 위치를 좌표로 매핑하여 웹에서의 시각화를 위해 HTML을 사용하여 추정된 좌표 정보를 동적으로 업데이트한다. 사용자는 웹 브라우저를 통해 추정된 위치의 좌표를 실시간으로 확인할 수 있다.

3. 과제 설계 상세

3.1 데이터 셋

3.1.1 데이터 수집

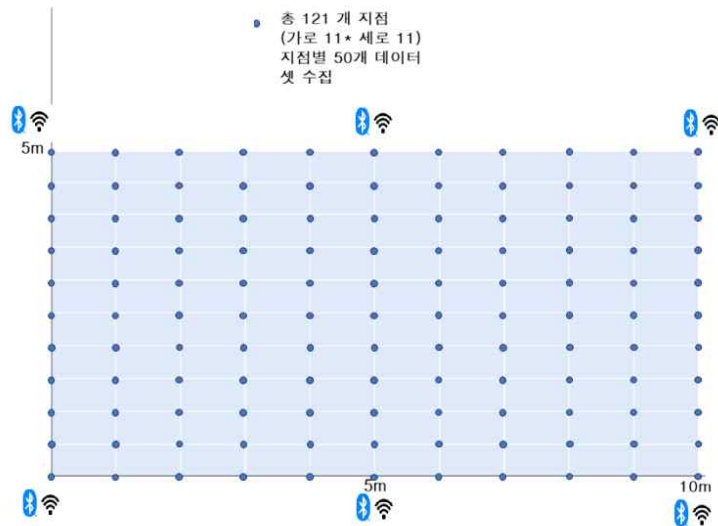


그림 9 위치별 초기데이터 수집 point

데이터 수집은 가로 10m, 세로 5m의 실내 공간에 5m 간격으로 총 6개의 스캐너를 배치하여 수집한다. 각 스캐너의 위치는(x, y)= (0,0), (5,0), (10,0), (0,5), (5,5), (10,5)이다. 이 스캐너들은 Lolin d32 보드를 이용하여 Beacon에서 발신되는 BLE 신호의 RSSI 값을 수집한다. 수집한 RSSI 값은 필터링을 거쳐 수집한다.

총 121개의 지점에서 각각 6개의 스캐너로부터 수집된 RSSI 값을 기록한다. 각 지점에서 수집된 데이터는 x와 y 좌표별로 6개의 스캐너에서 수집한 RSSI 값을 포함한 벡터로 구성되어있다. 따라서 총 121개의 지점에 대해 6개의 스캐너에서 수집한 RSSI 값의 벡터를 50개씩 수집한다.

value_1	value_2	value_3	value_4	value_5	value_6	x	y
-43.91	-63.67	-43.91	-50.71	-43.91	-26.45	10	5
-43.91	-65.89	-64.79	-65.89	-60.93	-31.42	10	5
-43.91	-65.89	-64.79	-70.16	-60.93	-31.5	10	5
-43.91	-65.89	-64.79	-71.3	-60.93	-31.32	10	5
-43.91	-65.89	-68.17	-71.3	-60.93	-31.18	10	5
-43.91	-66.91	-66.64	-72.23	-62.59	-31.67	10	5
-43.91	-66.91	-66.26	-73.62	-62.59	-32.28	10	5
-58.23	-67.02	-66.98	-73.62	-64.78	-31.95	10	5
-63.89	-67.02	-68.33	-73.62	-65.31	-32.02	10	5
-63.89	-67.76	-68.33	-73.92	-65.31	-32.46	10	5
-65.57	-67.76	-67.88	-74.16	-65.31	-32.92	10	5
-65.57	-67.76	-68.92	-74.16	-68.8	-32.87	10	5
-66.62	-67.95	-68.92	-74.35	-68.62	-32.69	10	5
-68.48	-68.59	-69.28	-74.35	-68.62	-32.65	10	5
-68.48	-68.59	-70.14	-74.35	-68.62	-32.94	10	5
-68.84	-68.69	-70.14	-75.47	-70.6	-32.81	10	5
-68.84	-68.69	-69.6	-75.47	-70.6	-33.07	10	5
-69.99	-68.65	-69.84	-75.68	-70.6	-32.95	10	5
-69.99	-68.65	-69.84	-76.14	-70.25	-33.14	10	5
-70.89	-69.05	-69.45	-76.14	-70.25	-33.35	10	5
-70.9	-69.05	-69.11	-76.2	-69.89	-33.3	10	5
-70.92	-69.05	-68.8	-76.25	-69.75	-33.19	10	5

그림 10 비콘 번호와 각 위치 별 RSSI 벡터 데이터셋

이 수집된 데이터는 데이터베이스에 저장하며, 이를 효과적으로 활용하기 위해 데이터를 전처리하고 CSV 파일로 변환한다. 전처리된 데이터는 과제에서 결과를 분석하고 모델을 학습하는 데 사용된다.

3.1.2 RSSI filtering

초기 RSSI 데이터는 노이즈와 불안정성이 포함되어 있어 정확한 분석을 위해 필터링이 필요하다. Raw 데이터와 Kalman Filter를 적용한 후, 3m 지점의 수집된 각각의 50개 샘플 데이터를 표준편차를 계산한다. 표준편차를 이용해 두 데이터셋 간의 안정성과 신뢰성을 비교한다. 이렇게 정제된 데이터는 실험의 신뢰성을 높이고 더 정확한 결과를 얻을 수 있게 한다.

3.2 모델구성

3.2.1 GPR & kernel 비교와 베이지안 최적화(Bayesian Optimization)

베이지안 최적화를 활용하여 각 커널의 하이퍼 파라미터를 조합하고 최적의 성능을 낼 수 있는 하이퍼 파라미터 조합을 찾아 낸 후, Gaussian Process Regression(GPR) 에서 사용할 kernel을 비교하고 최적의 커널을 선정한다. RBF(Radial Basis Function) kernel, Matern kernel, White kernel, Quadratic kernel을 각 커널별로 에러의 Root Mean Square Error (RMSE)를 계산하여 성능을 평가한다.

이를 통해 각 커널이 주어진 데이터에 어떻게 적합한지 비교하고, 최종적으로 RMSE가 가장 낮은 커널을 선택하여 최적의 모델을 결정한다.

3.2.2 GAN을 통한 데이터 증강(data augmentation with GAN)

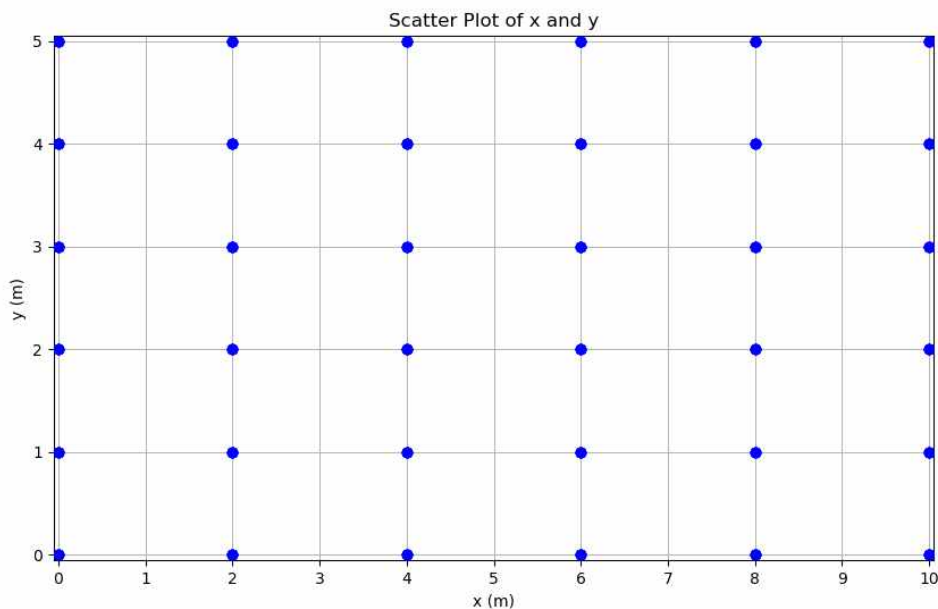


그림 11 데이터 확장 전 수집(Limited Data)

GAN을 통한 fake data 생성은 다양성을 확보하여 모델의 실제 시나리오 대응 능력을 강화하고, 예측 모델의 성능을 향상 시키며, 실제 데이터 수집에 비해 비용과 시간을 절약할 수 있는 효과적인 방법이다. 본 과제에서는 원 데이터와 제한된 데이터(limited)에서 GAN을 사용하여 증강된 데이터를 활용하여 GPR 모

델을 훈련한다. 데이터 증강(data augmentation)은 다음의 과정을 거쳐 생성된다.

먼저, 원 데이터와 제한된 데이터로부터 수집된 RSSI 벡터로 지도학습 가우시안 프로세스 회귀 (GPR) 모델을 훈련한다. 다음으로, GAN을 활용하여 fake data를 생성합니다. 이러한 fake data는 훈련된 GPR 모델을 기반으로 생성된다.

fake data 생성 후, 훈련된 GPR 모델을 활용하여 fake data에 대한 예측된 위치 정보를 예측한다. 그 다음, 생성된 fake data 중에서 실제 데이터와 유사성 비율을 기준으로 선택된 현실적인 fake data를 선별한다.

마지막으로, 선별된 fake data와 실제 데이터를 활용하여 위치 결정을 위한 혼합 가우시안 프로세스 회귀 (GPR) 모델을 훈련한다. 이렇게 훈련된 모델은 실제 환경에서 정확한 위치 추정을 수행할 수 있게 됩니다. 이를 통해 우리는 다양한 데이터를 활용하여 모델의 성능을 향상시키고, 실제 데이터 수집에 비해 비용과 시간을 절약할 수 있는 효과적인 방법을 구현하게 된다.

<실험 그룹 구성>

세 가지 실험 그룹을 구성하였다.

(Limited Data, Extensive Data, Limited Data+GAN.)

-Limited Data 그룹은 원래 수집된 제한된 데이터만 사용하여 모델을 훈련하고 평가한다.

-Extensive Data 그룹은 실제 데이터 수집 범위를 확장하여 데이터를 수집하고 모델을 훈련한다.

-Limited Data+GAN 그룹은 제한된 데이터에 GAN을 적용하여 생성된 fake data를 활용하여 모델을 훈련한다.

<평가 지표 선택>

실험 그룹 간의 성능 비교를 위해 MSE와 90% CDF를 선택하였다. MSE는 정확도를, 90% CDF는 모델의 예측 범위를 나타내는 지표로 사용된다.

3.3 웹서비스 구현

3.3.1 Zoning system

Zoning Visualization 은 사용자에게 현재 위치를 여러 영역 또는 "Zone"으로 구분하여 시각적으로 제공한다. 이를 위해 Python의 Flask 라이브러리와 pymysql 라이브러리를 사용하여 실시간 RSSI 값을 가져오고, 학습된 Gaussian Process Regression (GPR) 모델에 적용하여 얻은 XY 좌표를 Weighted k-Nearest Neighbors (WkNN) 알고리즘을 이용해 추정한다. 이때, Zoning을 위해 지정한 범위는 X 축에서 [0, 3.3, 6.6, 10], Y 축에서 [0, 2.5, 5]로 설정되어 있다. 이는 총 6개의 Zone으로 나뉜다.

실시간으로 수집한 RSSI 값을 Flask 서버를 통해 받아오고, 이를 pymysql 라이브러리를 이용하여 데이터베이스에서 적절한 형식으로 가져온다. 이후, 가져온 데이터는 모델의 입력 형식에 맞게 전처리된다.

<위치 추정과 Zoning>

전처리된 데이터는 학습된 GPR 모델에 입력으로 제공되어 XY 좌표로 변환된다. 이후, 변환된 XY 좌표는 WkNN 알고리즘을 통해 실시간으로 추정된 사용자 위치를 나타내게 된다. 추정된 XY 좌표는 정해진 Zone으로 매핑되어 사용자에게 시각적으로 제공된다.

<Zoning Accuracy>

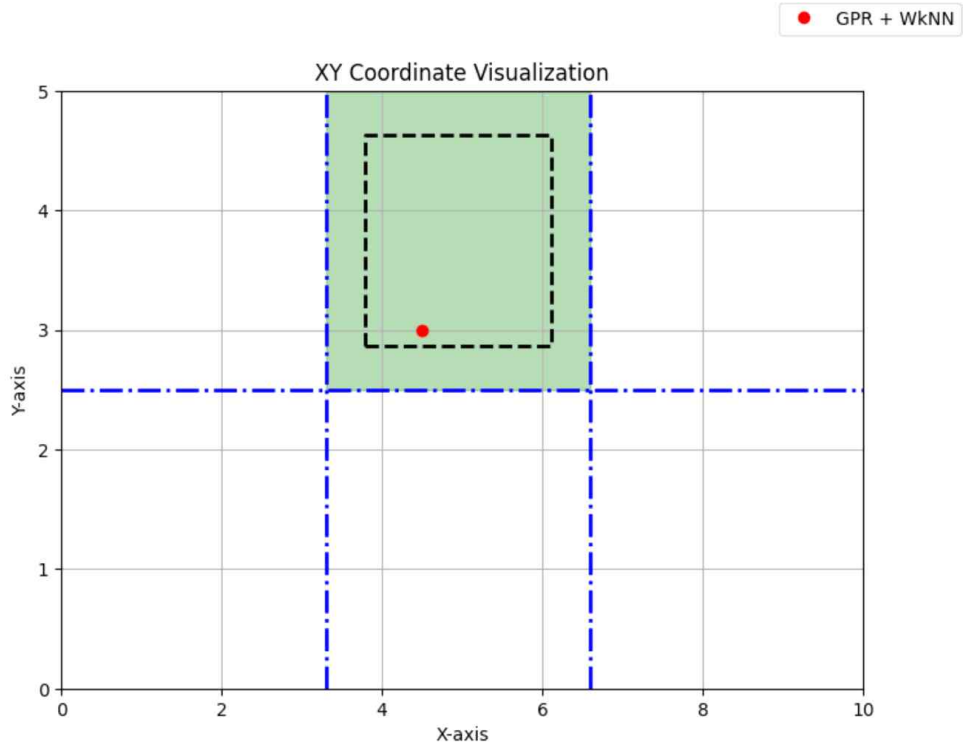


그림 12 zone inner/edge 경계

예측된 좌표를 6개의 블록으로 나누어 각 블록에 속하는지 확인하고 정확도를 계산한다. 총 214개의 지점에서, 실제 좌표와 zoning을 비교하여 세 가지 정확도를 계산하며, 내부와 가장자리는 zone의 총 넓이를 기준으로 50%가 되는 지점을 경계로 한다. 이는 다음과 같다:

전체 정확도(Total Accuracy)

예측된 좌표가 실제 Zone과 일치하는 비율로 전체 데이터 셋에 대한 정확도.

Zone 내부 정확도(Inner Accuracy)

예측된 좌표가 실제 Zone 내부에 속하는지 확인하여 정확도를 계산.

Zone 가장자리 정확도(Edge Accuracy)

예측된 좌표가 실제 블록의 가장자리에 속하는지 확인하여 정확도를 계산.

이러한 정확도 측정 결과는 zoning 시스템의 성능을 평가하고, zone의 경계부분과 안쪽 부분의 정확도를 구분할 수 있다.

<시각화 및 사용자 인터페이스>

시각화는 추정된 XY 좌표와 해당하는 Zone을 매핑하여 웹 페이지 상에서 동적으로 업데이트된다. 사용자는 화면 상에서 이동체의 위치가 어떤 Zone에 속하는지 확인할 수 있다.

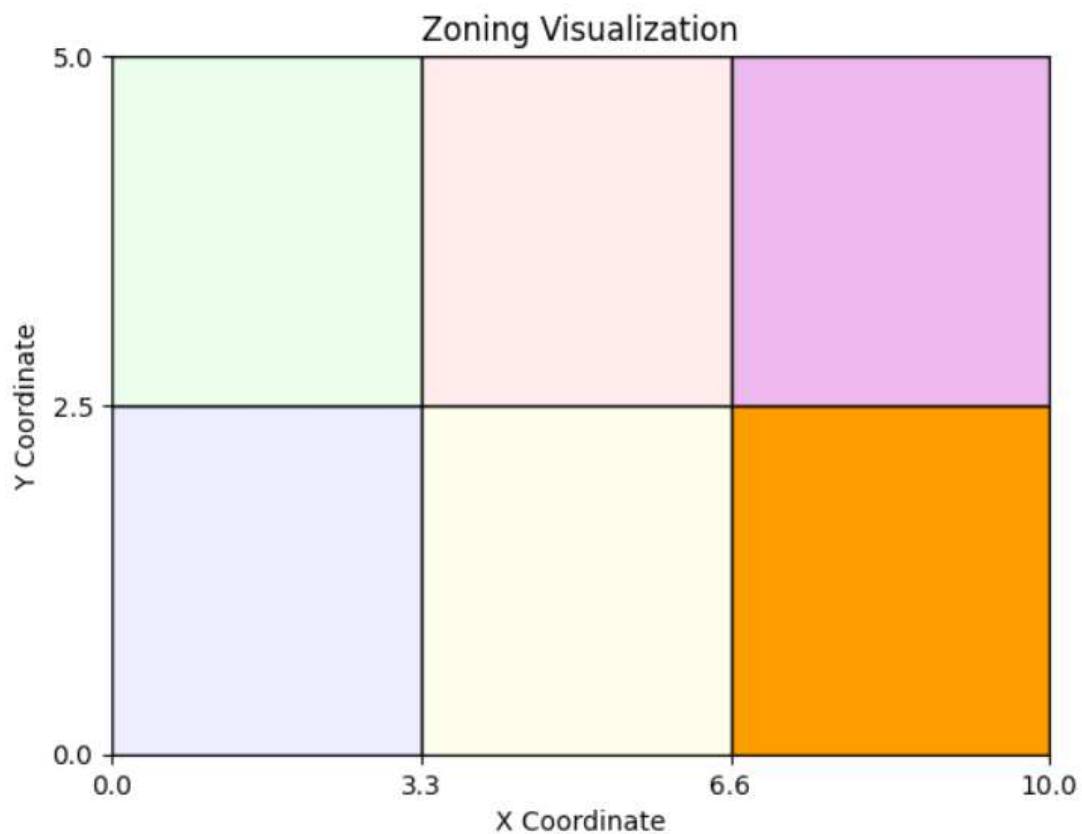


그림 13 6개의 Zone / Visualization

3.3.2 Trajectory reconstruction

다음 3가지의 경로를 만들어 수집된 RSSI값을 바탕으로 예측된 위치를 기반으로 경로를 추정해본다. Estimated Path (추정된 경로), Reconstructed Path (재구성된 경로)를 시각적으로 비교한다.

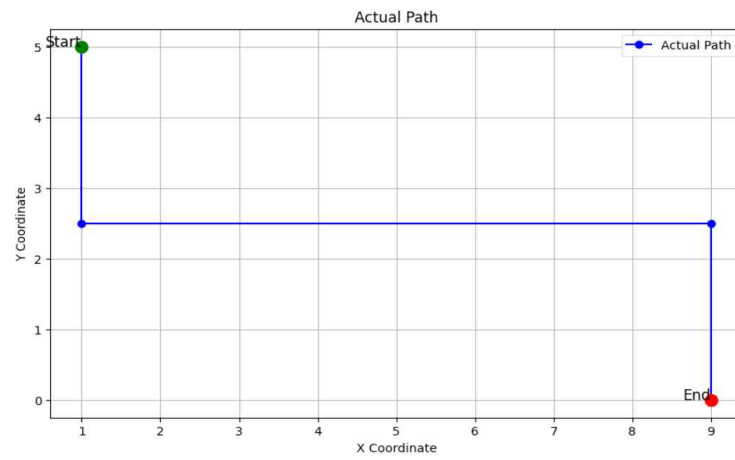


그림 14 경로_1

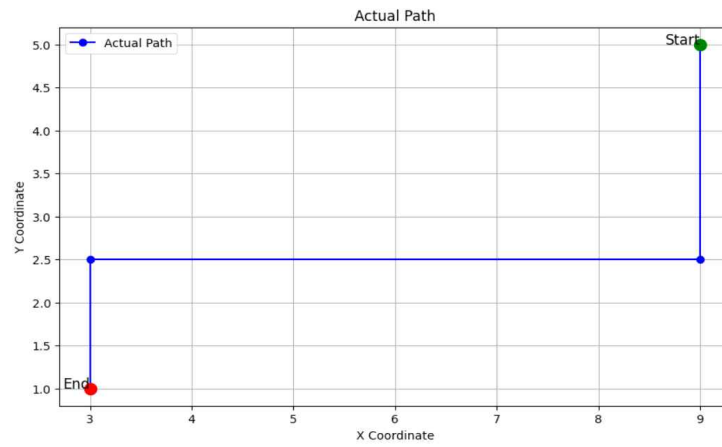


그림 15 경로_2

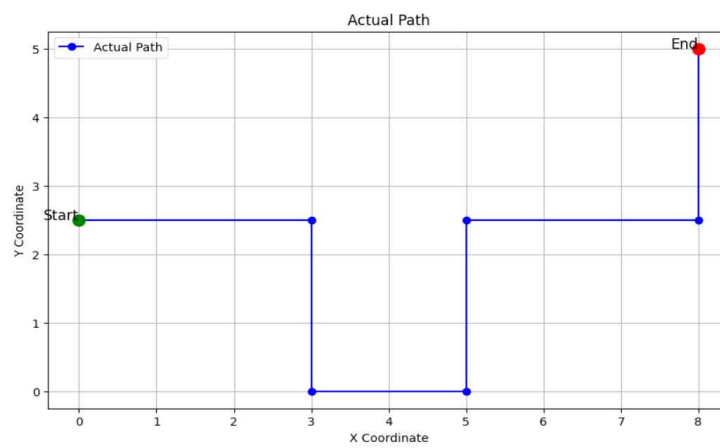


그림 16 경로_3

Estimated Path는 주어진 RSSI 값들을 기반으로 미리 학습된 모델을 사용하여 예측된 경로를 나타낸다. 이 경로는 단순히 RSSI 값과 위치 간의 관계를 기반으로 예측된 것이며, 주어진 데이터를 정확하게 반영하고 있을 수 있다.

Reconstructed Path (재구성된 경로)는 RSSI 값들의 신뢰성을 평가하고, 추정된 위치를 가중 평균을 통해 최적화한다. 이는 다음의 과정을 거쳐 생성된다.

<데이터 신뢰성 평가 (Reliability Assessment)>

먼저, 주어진 RSSI 값들과 예측된 위치 간의 유사성을 평가하여 신뢰성 있는 포인트들을 선별합니다. threshold로 정의된 임계값을 사용하여, RSSI 값과 예측 위치 간의 거리를 계산하고 이 임계값 이하의 거리에 있는 포인트를 신뢰성 있는 포인트로 선택합니다.

<가중치 부여 및 가중평균 계산 (Weighted Averaging)>

선별된 신뢰성 있는 포인트들을 기반으로 주변 예측된 위치들에 대한 가중치를 계산한다. 각 포인트의 가중치는 거리에 반비례하여 계산되며, 더 가까운 포인트일수록 더 큰 가중치를 받는다. 이렇게 계산된 가중치를 사용하여 예측된 위치들의 가중평균을 계산한다.

<경로 보간 (Path Interpolation)>

가중 평균을 통해 얻은 위치들을 기반으로 경로를 보간한다. 이 보간은 선형 보간을 사용하여 이루어진다. 먼저, 신뢰성 있는 포인트들 사이의 직선 경로를 생성하고, 이를 통해 최적의 경로를 재구성한다. 보간된 경로는 원래의 예측된 위치들을 보다 부드럽게 이어주어, 경로를 더 정확하게 나타낸다.

4. 과제 수행 내용 및 결과

4.1 Data filtering

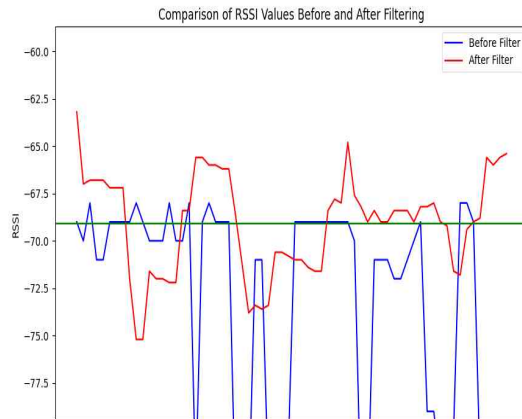


그림 17 Before vs After filtering

state	표준편차
Before filtering	5.9889722374 99391
After filtering	2.6609723585 136873

표 1 표준편차 (before vs after filtering)

칼만 필터(kalman filter)를 적용한 결과, 초기에 5.99였던 표준편차가 2.66 으로 현저히 감소하였다. 이를 통해 데이터의 불안정성이 크게 줄어들었으며, 이를 시각적으로 나타내기 위해 도표와 표를 사용하였다. 기존의 Raw 데이터는 노이즈가 많아 사용이 어려웠지만, 칼만 필터를 통해 데이터를 정제함으로써 보다 정확한 분석이 가능하다. 이렇게 필터링된 데이터를 기반으로 한 결과해석은 실험의 신뢰성을 높이고, 보다 정확하고 신뢰할 수 있는 결과를 도출할 수 있다.

4.2 커널 선택

무선 신호의 거리에 따른 감쇠 현상을 고려할 때, 더 멀리 떨어진 x 좌표에서의 노이즈 수준이 더 높을 수 있다는 점에서 전체적인 RMSE값이 x 좌표에서 더욱 크게 나타나고 있다.

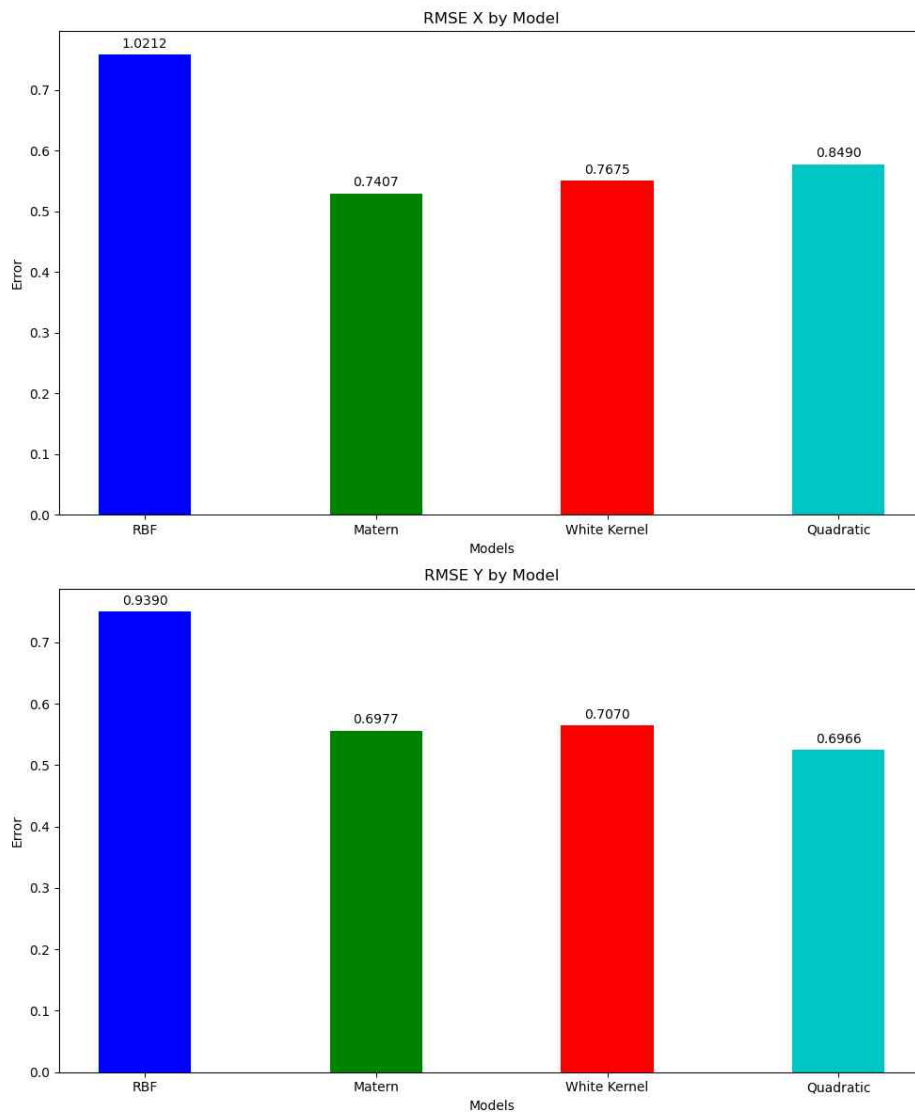


그림 18 좌표별 커널 성능비교(RMSE)

4.2.1 Model x

x 좌표 모델에서는 Matern 커널이 10m 방향에서의 데이터 패턴을 부드럽게 모델링하여 가장 작은 RMSE 값(RMSE: 0.7407)을 보였다. 이는 Matern 커널이 불규칙한 환경에서도 안정적인 예측을 제공함을 나타낸다고 해석된다.

4.2.2 Model y

y 좌표 모델에서는 Quadratic 커널이 5m 방향에서의 비선형 데이터 패턴을 잘 파악하여 가장 작은 RMSE 값을 보였다. 이는 Quadratic 커널이 비선형적인 데이터 특성을 잘 모델링하면서 노이즈를 효과적으로 처리했음을 나타내고 있다.

4.3 데이터 증강(data augmentation)

4.3.1 Generated fake data/ Limited data + selected pseudo-labeled data scatter plot

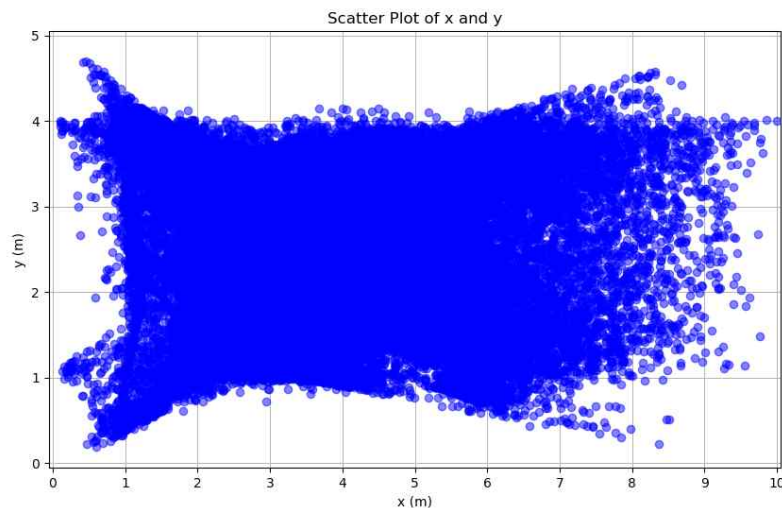


그림 19 Generated fake data

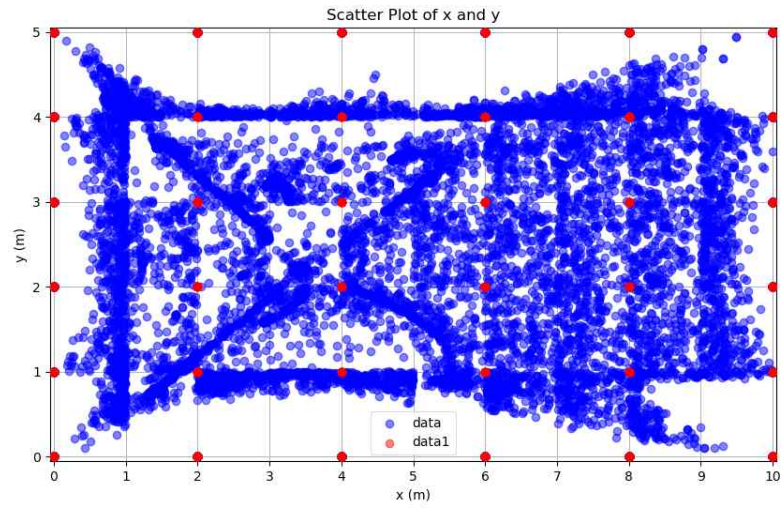


그림 20 Limited data + selected pseudo -labeled data scatter plot

4.3.2 정확도 비교

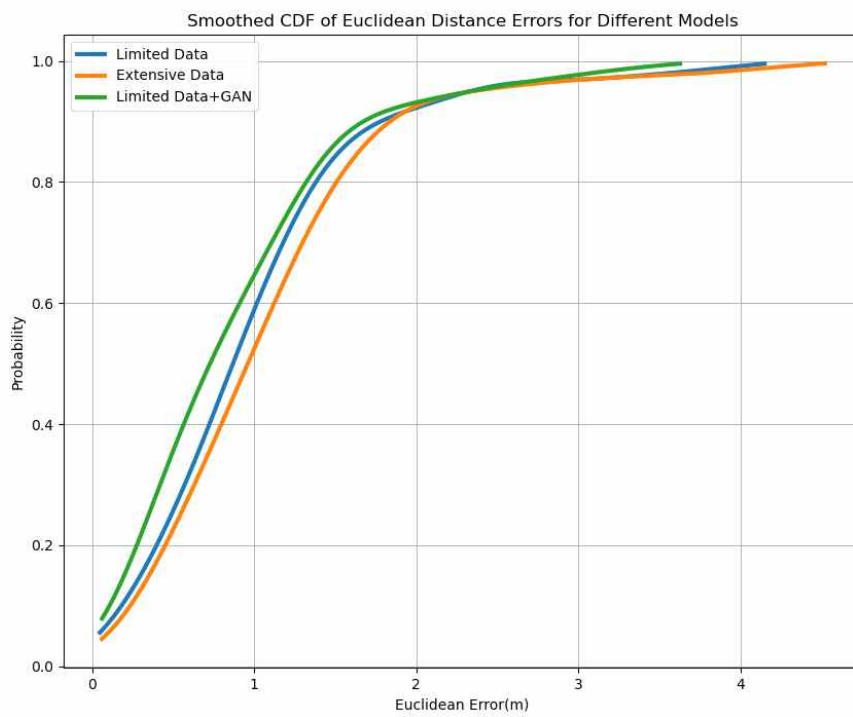


그림 21 Smoothed CDF of Euclidean Distance Errors for Different Models

	Limited Data	Extensive Data	Limited Data+GAN
MSE	1.4576	1.6611	1.2061
90% CDF	1.6569	1.7148	1.5534

표 2 MSE, 90% CDF (limited vs extensive vs limited+GAN)

<Mean Squared Error (MSE)>

MSE는 모델의 예측값과 실제 값 간의 평균적인 제곱 오차를 나타낸다. MSE 값이 낮을수록 모델의 예측이 실제 값과 가깝다고 볼 수 있다. Limited Data+GAN 조건에서 가장 낮은 MSE 값인 1.2061을 보이므로, fake data를 활용한 경우에 모델의 예측 정확도가 가장 높았다고 할 수 있다.

<90% Cumulative Distribution Function (90% CDF)>

90% CDF는 예측 값이 실제 값의 90%를 커버하는 범위를 나타낸다. 이 값이 높을수록 모델이 더 넓은 범위에서 정확한 예측을 수행한다는 의미이다. Limited Data+GAN 조건에서 90% CDF 값이 1.5534로, 모델이 실제 값의 90%를 커버하는 범위가 가장 넓다고 할 수 있다.

결과적으로, fake data를 활용한 Limited Data+GAN 조건이 모델의 정확도와 예측 범위 측면에서 가장 우수한 결과를 보였다. 이는 GAN을 사용하여 데이터 다양성을 향상 시키고, 모델의 성능을 향상 시키는 데 효과적이었다고 해석할 수 있다.

4.4 Zoning System

4.4.1 Zoning visualization & accuracy

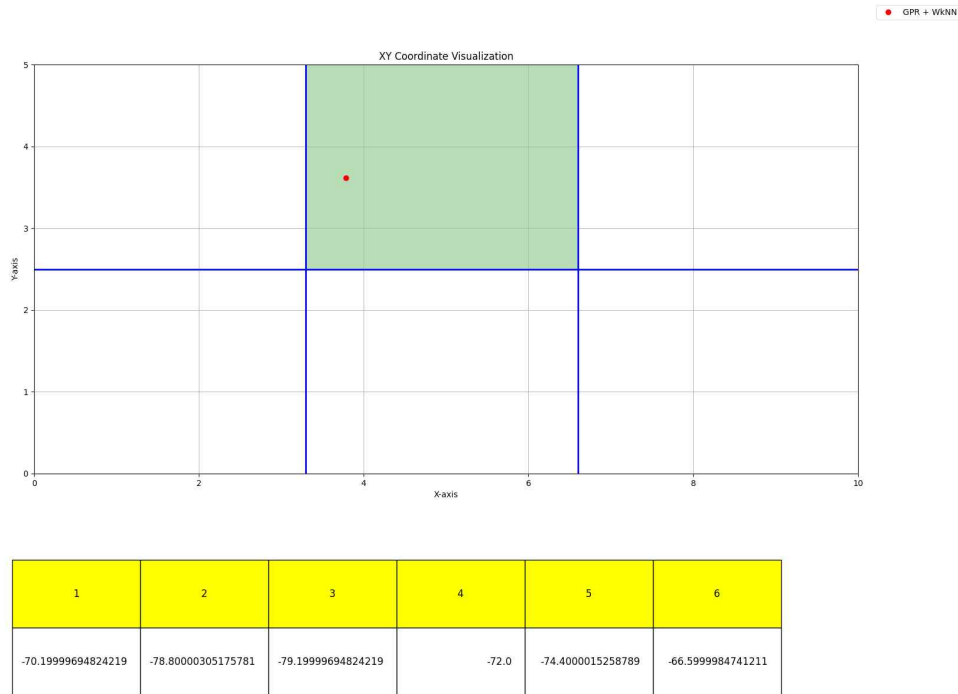


그림 22 Zoning Visualization

Zoning Visualization은 실시간으로 수집된 RSSI 값을 기반으로 사용자의 현재 위치를 여러 영역 또는 "Zone"으로 구분하여 시각적으로 제공한다. 이를 통해 사용자는 웹 페이지 상에서 이동체의 위치가 어떤 Zone에 속하는지 확인할 수 있다.

Category	Accuracy (%)
Total	71.96%
Inner	100.00%
Edge	55.00%

표 3 Accuracy

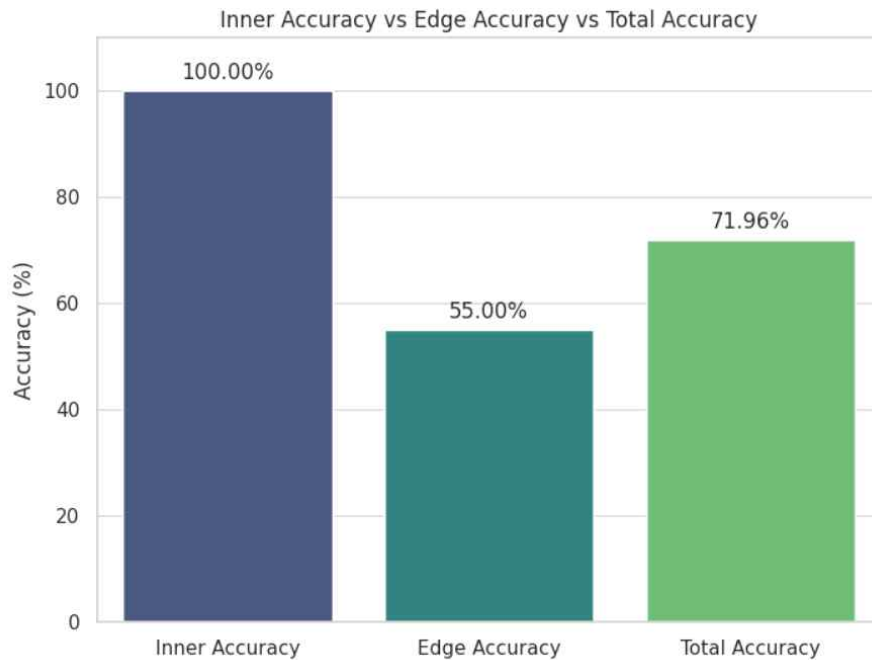


그림 23 accuracy of zoning

예측된 좌표를 6개의 블록으로 나누어 각 블록에 속하는지 확인하고 정확도를 계산한다. 총 214개의 지점에서, 실제 좌표와 zoning을 비교하여 세 가지 정확도를 계산하며, 내부와 가장자리는 zone의 총 넓이를 기준으로 50%가 되는 지점을 경계로 한다.

Total Accuracy (전체 정확도): 71.96%

전체 정확도는 시스템이 예측한 위치가 실제 위치와 얼마나 일치하는지를 나타낸다. 이 모델은 주어진 데이터셋에서 약 72%의 정확도로 위치를 예측할 수 있었다. 즉, 시스템이 예측한 위치 중 약 72%가 실제 위치와 일치한다고 볼 수 있다.

Inner Accuracy (Zone 내부 정확도): 100.00%

Zone 내부 정확도는 시스템이 예측한 위치가 실제 블록 내부에 속하는 경우의 정확도를 나타낸다. 이 모델은 실제 Zone 내부에 속하는 경우에 대해서 100%의 정확도를 보였다. 즉, 시스템이 블록 내부에 위치한 데이터 포인트를 정확하게 예측할 수 있다.

Edge Accuracy (Zone 가장자리 정확도): 55.00%

Zone 가장자리 정확도는 시스템이 예측한 위치가 실제 블록의 가장자리에 속하는 경우의 정확도를 나타낸다. 이 모델은 실제 Zone 가장자리 부분에 대해서

55%의 정확도를 보였다. 즉, 시스템이 블록 가장자리 부분의 위치를 예측하는데에는 55%의 정확도를 가지고 있다.

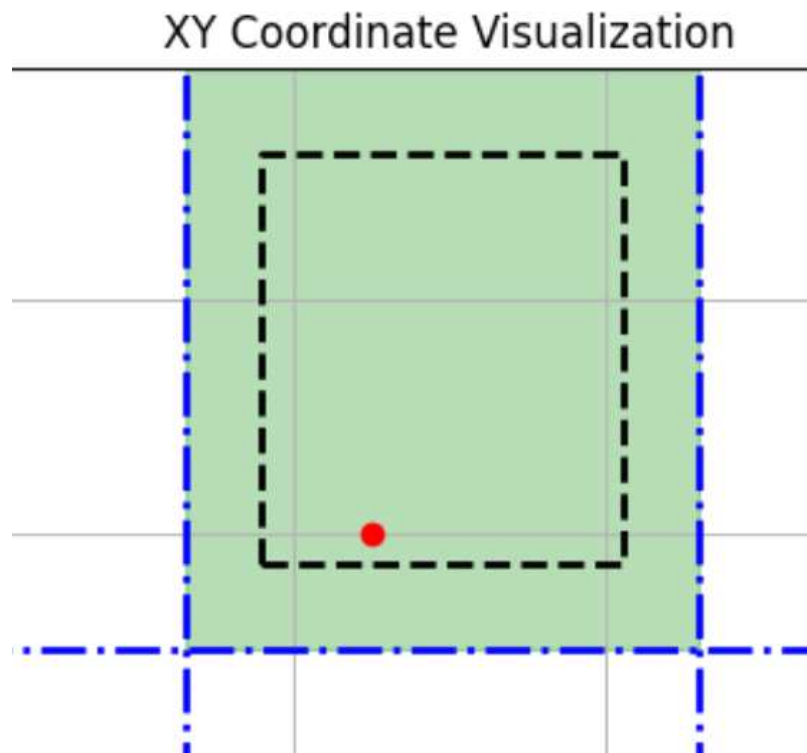


그림 24 edge/ inner boundary

이러한 결과는 전체적으로 시스템이 블록 내부에 위치한 데이터를 매우 정확하게 예측할 수 있지만, Zone 가장자리 부분에 대해서는 정확도가 낮다는 것을 나타낸다.

4.4.2 Trajectory reconstruction

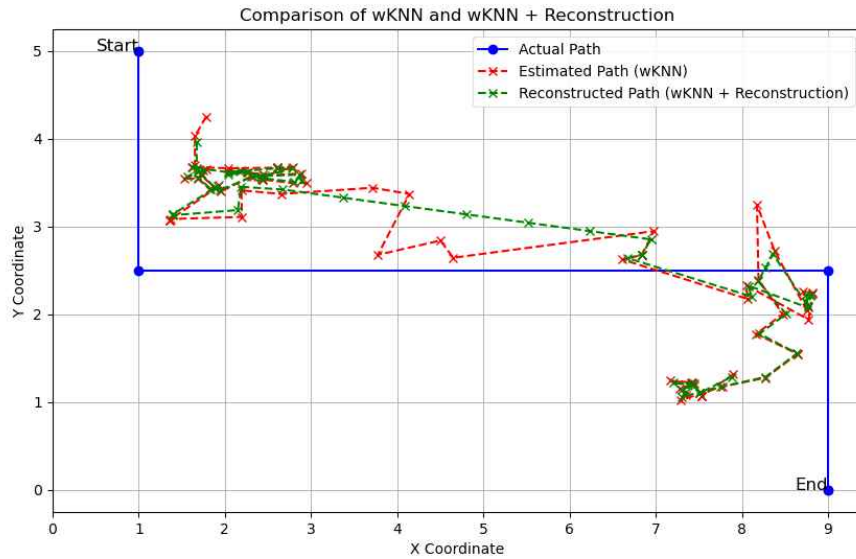


그림 25 Comparison of wKNN & wKNN+Reconstruction 경로_1

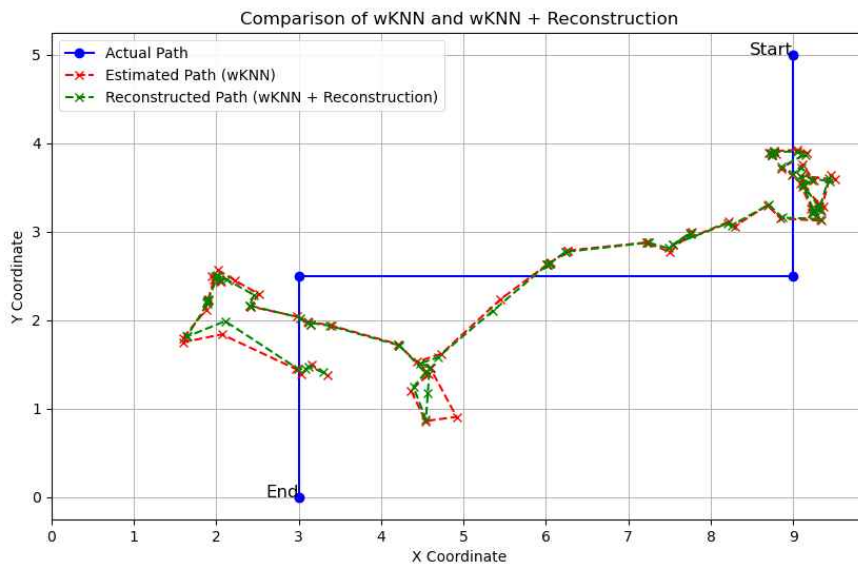


그림 26 Comparison of wKNN & wKNN+Reconstruction 경로_2

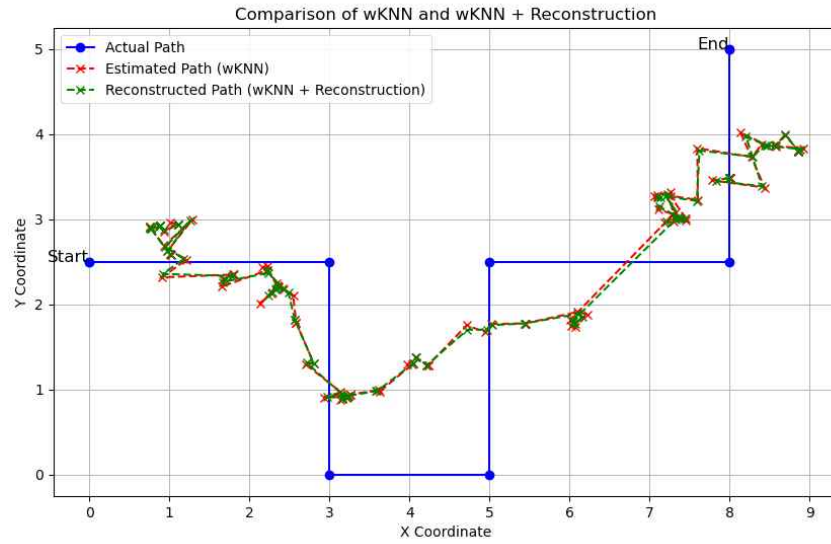


그림 27 Comparison of wKNN & wKNN+Reconstruction 경로_3

세 경로의 경우, 시각적으로 보았을 때 Reconstructed Path는 실제 경로와 매우 유사하게 나타난다. 부드러운 곡선으로 연결되어 있어 시각적으로 자연스럽다. 그러나, 정확한 경로 추적에 있어서는 시각적으로 정확도가 부족하다.

5. 결론 및 향후 연구 방향

5.1 결론

본 과제에서는 무선 신호 데이터를 기반으로 한 위치 추정 시스템을 개발하고 성능을 향상시키기 위한 다양한 기술적 접근 방법을 탐구하였다. 주요 결과와 결론은 다음과 같다:

<데이터 필터링 및 정제>

칼만 필터를 통한 데이터 필터링은 초기 표준편차를 현저하게 감소시켜 데이터 불안정성을 줄였다. 이로써 정확한 분석이 가능하게 되었고, 신뢰성 있는 결과 도출을 가능케 했다.

<커널 선택 및 성능 비교>

각각의 좌표 모델에 대해 적절한 커널 선택은 모델의 정확도에 큰 영향을 미쳤다. Matern 커널과 Quadratic 커널은 각각 특정 방향에서의 데이터 패턴을 잘 파악하여 최소의 RMSE 값을 보였다.

<데이터 증강 기법 적용>

Fake 데이터를 활용한 Limited Data+GAN 조건은 모델의 정확도를 획기적으로 향상 시켰다. MSE 값과 90% CDF 값 모두에서 우수한 성능을 보였다. 이는 적은 데이터로 충분히 모델의 예측 정확도와 범위를 확대시킬 수 있음을 보여준다.

<Zoning 시스템 구현>

무선 신호 데이터를 기반으로 한 Zoning 시스템은 사용자의 위치를 정확하게 파악하여 여러 Zone으로 분류하고 시각적으로 표현하는 데 성공하였다.

종합적으로, 이 연구는 무선 신호를 활용한 위치 추정 시스템의 정확성과 신뢰성을 향상시키기 위한 다양한 방법론을 탐구하였고, Fake 데이터와 GAN을 활용한 데이터 증강 기법이 모델 성능 향상에 효과적임을 입증하였다. 또한, Zoning 시스템은 이동체의 위치 추적을 시각적으로 표현함으로써 실내 물류창고에서 이동 및 근로자의 움직임을 효율적으로 관리하고 추적을 가능케 해 운영 최적화와 생산성 향상에 도움이 될 것이다.

5.2 산학협력 멘토링

산학협력 멘토링을 통해 다양한 측면에서 졸업과제를 발전시킬 수 있는 중요한 지점들을 확인하였다. 멘토링 세션은 간단한 발표 형식으로 진행되었다.

먼저, 물류창고에서의 데이터 정확도를 높이기 위해서는 환경적인 요인들을 먼저 정형화해야 한다는 점을 확인할 수 있었다. 환경적인 변인들은 물류 시스템에서 예측할 때 큰 영향을 미칠 수 있으며, 이를 고려하지 않으면 정확한 예측이 어려울 수 있으므로 정확도보다는 다른 쪽에 방향성을 제시받았다.

확보된 데이터를 효과적으로 표현하는 것, 데이터는 풍부하지만 이를 효과적으로 시각화하고 해석하는 것이 프로젝트의 가치를 높일 수 있다는 점을 확인할 수 있었다.

멘토링 동안 유사한 서비스 사례(ex. 디지털 트윈 등)를 보았으며, 이를 참고하여 연구 방향을 잡는 것이 중요하다는 점도 배웠다. 다양한 참고 자료를 바탕으로 시각화 부분의 방향성을 잡고 조금 더 발전시킬 수 있었다.

5.3 향후 연구 방향

본 과제에서는 무선 신호 데이터를 기반으로 한 위치 추정과 Reconstructed Path의 시각적 표현을 통한 이동체 추적을 성공적으로 구현하였다. 그러나, 정확한 경로 추적에서 아직까지는 한계가 있음을 확인하였다. 이를 극복하고 더 나은 결과를 얻기 위해 다음과 같은 향후 연구 방향을 고려할 수 있다.

<다중 센서 통합>

여러 종류의 센서 데이터를 통합하여 더 다양한 정보를 활용하는 것이 가능하다. 예를 들어, 가속도계 및 자이로스코프와 같은 센서 데이터를 활용하여 움직임의 방향과 속도등을 활용해 더욱 정밀한 알고리즘을 구현할 수 있다.

<신경망 기반 추정 시스템 도입>

딥러닝 기술을 활용한 신경망을 구성하여 위치 추정을 시도할 수 있다.

<환경 모델링과 향상된 경로 예측>

물체의 이동 패턴 및 환경 요인을 더 정확하게 모델링하고, 미래의 이동 경로를 더 정확하게 예측하는 모델을 개발할 수 있다. 환경의 변화에 빠르게 대응하면서도 정확한 예측을 수행하는 모델을 구성하는 것이 보다 정확한 위치 추정을 가능케 할 것이다.

<실시간 보정 및 업데이트 기능 강화>

이동체가 움직이는 동안 실시간으로 모델을 보정하고 업데이트하는 기능을 강화함으로써 더욱 정확한 경로 추적이 가능해질 것이다.

<다중 경로 간섭 관리 및 모델링>

무선 신호는 다수의 경로를 통해 수신될 수 있다. 향후 연구에서는 다중 경로 간섭 현상을 더 정확하게 모델링하고 관리하는 기술을 개발하면 각 경로의 신호 세기와 방향을 고려하여 정확한 위치를 추정할 수 있을 것이다.

<지능적인 신호 감지와 회피 알고리즘의 개발>

주변 환경의 무선 신호를 실시간으로 감지하고, 간섭이 발생하는 경우에 자동으로 신호를 조정하는 지능적인 알고리즘을 하여, 동적으로 주파수 및 방향을 조절하여 간섭을 회피하고 정확한 위치 추정에 영향을 미치지 않도록 할 수 있을 것이다.

이러한 향후 연구 방향들은 정확한 위치 추정과 경로 추적 기술을 더욱 발전시켜 물류창고나 다른 실내 환경에서의 자동화 및 효율적인 운영을 지원할 수 있는 가능성을 제시한다.

6. 개발 일정 및 역할 분담

6.1 개발 일정

6월				7월				8월				10월	
1주차	2주차	3주차	4주차	1주차	2주차	3주차	4주차	1주차	2주차	3주차	4주차	1주차	2주차
데이터 셋 확보													
기본 모델 작성													
모델 최적화													
중간보고서 작성													
예외 처리 및 모델 수정													
최종 보고서 작성													

표 4 개발일정

6.2 역할 분담

이름	역할
이정현	RSSI 데이터 수집 환경구축 및 수집 모델 결과분석 보고서, 포스터 작성 및 발표
김지명	RSSI 데이터 수집 환경구축 및 수집 GPR 학습 모델 구성 모델 최적화 모델 결과분석
강중현	RSSI 데이터 수집 환경구축 및 수집 학습에 필요한 MySQL 데이터베이스 구축 모델 최적화 실시간 위치 정보 웹서비스 구현

표 5 역할 분담

7. 참고 문헌

[1]Bluetooth Technology Overview . (2023, Oct 13).
Available:<https://www.bluetooth.com/learn-about-bluetooth/tech-overview/>

[2]W. Njima, M. Chafii, A. Chorti, R. M. Shubair and H. V. Poor, "Indoor

Localization Using Data Augmentation via Selective Generative Adversarial Networks," in *IEEE Access*, vol. 9, pp. 98337–98347, 2021, doi: 10.1109/ACCESS.2021.3095546.

[3]D. D. Nguyen and M. Thuy Le, "Enhanced Indoor Localization Based BLE Using Gaussian Process Regression and Improved Weighted kNN," in *IEEE Access*, vol. 9, pp. 143795–143806, 2021, doi: 10.1109/ACCESS.2021.3122011.