

TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI
TRƯỜNG ĐIỆN – ĐIỆN TỬ

ĐỒ ÁN II

**ĐỀ XUẤT THUẬT TOÁN CẢI THIẾN ĐỘ CHÍNH XÁC VỊ TRÍ TRONG ĐỊNH
VỊ TRONG NHÀ SỬ DỤNG CÔNG NGHỆ WIFI VÀ BLUETOOTH**

NGUYỄN HUY HOÀNG

Hoang.nh191855@sis.hust.edu.vn

Ngành Kỹ thuật điều khiển và tự động hóa
Chuyên ngành Kỹ thuật đo và tin học công nghiệp

Giảng viên hướng dẫn: TS. Nguyễn Đại Dương

Chữ ký của GVHD

Bộ môn: Tự động hóa

Viện: Điện

HÀ NỘI, 3/2023

Lời cảm ơn

Lời đầu tiên em xin gửi lời cảm ơn đến thầy Nguyễn Đại Dương người hướng dẫn trực tiếp em thực hiện đề tài này. Thầy đã có những định hướng và trao đổi với em trong suốt thời gian em thực hiện đồ án II của mình. Thầy đã hướng dẫn chỉ dạy em hoàn thành đề tài cũng như báo cáo học phần đồ án II. Vì hiểu biết hạn hẹp và thời gian hạn chế nên còn nhiều sai sót và em mong nhận được những góp ý, sửa đổi từ thầy để đề tài của em được hoàn thiện hơn.

Em xin chân thành cảm ơn thầy!

Tóm tắt nội dung đồ án

Báo cáo phục vụ yêu cầu học phần Đồ án II, kì 20221. Mục tiêu của báo cáo là nghiên cứu và đề xuất các thuật toán để cải thiện độ chính xác của sai số vị trí sử dụng công nghệ định vị trong nhà Wifi và Bluetooth. Các phương pháp và thuật toán đề xuất có thể kể tới như: KNN, WKNN, P&S-WKNN, kết hợp WKNN với các thuật toán nội suy như SVM, Gradient boosting, Random Forest, nội suy KNN.

MỤC LỤC

DANH MỤC HÌNH ẢNH.....	4
DANH MỤC HÌNH VẼ, BẢNG BIỂU.....	5
CHƯƠNG 1: TÌM HIỂU CHUNG VỀ HỆ THỐNG ĐỊNH VỊ TRONG NHÀ	6
1.1 Bài toán định vị trong nhà.....	6
1.2 Tổng quan về công nghệ WiFi và Bluetooth	7
1.2.1 Tổng quan về công nghệ định vị	7
1.2.2 Phương pháp công nghệ định vị	8
CHƯƠNG 2: ĐỀ XUẤT CÁC THUẬT TOÁN, PHƯƠNG PHÁP	11
2.1 Đánh giá một vài phương pháp định vị	11
2.2 Phương pháp định vị Fingerprint.....	13
2.2.1 Tổng quan về phương pháp Fingerprint	13
2.2.2 Các nhân tố ảnh hưởng tới phương pháp định vị Fingerprint	15
2.3 Đề xuất các thuật toán và phương pháp định vị	16
2.3.1 Thuật toán KNN truyền thống.....	17
2.3.2 Thuật toán WKNN.....	18
2.3.3 Thuật toán kết hợp giữa P-WKNN và S-WKNN	21
2.3.4 Các thuật toán nội suy	25
CHƯƠNG 3: KẾT QUẢ THỬ NGHIỆM ĐỘ CHÍNH XÁC	30
3.1 Đánh giá sai số tập dữ liệu thứ nhất	30
3.2 Đánh giá sai số tập dữ liệu thứ hai	32
3.3 Nhận xét.....	33
CHƯƠNG 4: KẾT LUẬN.....	35
4.1 Kết luận.....	35
4.2 Đề xuất các giải pháp khác	35
TÀI LIỆU THAM KHẢO	36

DANH MỤC HÌNH ẢNH

Hình 1. 1 Phương pháp Triangulation	9
Hình 1. 2 Cấu trúc của công nghệ định vị bằng WiFi	10
Hình 2. 1 Mô hình tổng quan của phương pháp Fingerprint.....	13
Hình 2. 2 Mô hình thuật toán sử dụng trong pha Offline	14
Hình 2. 3 Mô hình trong pha Online	15
Hình 2. 4 Hệ thống định vị theo phương pháp Fingerprint	16
Hình 2. 5 Phương pháp định vị WKNN theo BLE.....	19
Hình 2. 6 Framework của thuật toán	21
Hình 2. 7 Khoảng cách vật lí và khoảng cách không gian	22
Hình 2. 8 Mô hình thuật toán Random forest.....	26
Hình 2. 9 Mô hình thuật toán SVM.....	27
Hình 2. 10 Mô hình thuật toán Gradient Boosting	28
Hình 3. 1 Đối tượng thực hiện thứ nhất.....	30
Hình 3. 2 Đối tượng thực hiện thứ hai.....	32

DANH MỤC HÌNH VẼ, BẢNG BIỂU

Bảng 1. 1 Kết quả tính toán sai số theo lí thuyết.....	12
Bảng 3. 1 Sai số tọa độ	31
Bảng 3. 2 Sai số RSS.....	31
Bảng 3. 3 Sai số tọa độ	33

CHƯƠNG 1: TÌM HIỂU CHUNG VỀ HỆ THỐNG ĐỊNH VỊ TRONG NHÀ

1.1 Bài toán định vị trong nhà

Kể từ những ngày đầu tiên các công nghệ định vị bằng vệ tinh xuất hiện vào năm 60 của thế kỷ trước, giải pháp cho việc định vị đã liên tục có những biến chuyển lớn để đáp ứng nhu cầu của con người và xã hội. Việc biết được vị trí của con người, của vật luôn luôn là mối quan tâm hàng đầu. Trong lĩnh vực vận chuyển, vị trí của xe hàng, tàu thủy, máy bay ... , nếu có thể biết được, sẽ làm tăng hiệu quả vận chuyển, là yếu tố quan trọng cho bài toán logistic đang rất “hot” trong những năm gần đây. Trong lĩnh vực bán hàng ở các trung tâm, siêu thị, ... , việc biết được số lượng khách hàng lui tới các khu vực quầy bán hàng nào, thời gian khách hàng lưu lại quầy đó là bao lâu, sẽ là dữ liệu quan trọng liên quan tới các chiến dịch bán hàng. Trong các nhà xưởng, kho bãi, ... có thể biết được loại hàng hóa, mã sản phẩm nào đang nằm ở đâu hoặc nhân viên đang ở vị trí nào trong xưởng, sẽ làm tăng hiệu quả sản xuất, đưa quy trình sản xuất, xuất nhập kho bãi số hóa đúng như mong muốn của các công ty lớn hiện nay. Đó chỉ là một trong những bài toán và giải pháp trong một “bể” những nhu cầu của thị trường và xã hội trong thời kỳ số hóa đang chuyển mình cực mạnh.

Gốc rễ của những giải pháp đó đều quy về một điểm đó là biết được vị trí của vật thể. Để có thể biết được yếu tố nghe tưởng chừng đơn giản đó, các nhà khoa học đã đưa ra rất nhiều công nghệ khác nhau, với độ hiệu quả khác nhau và phù hợp cho các bài toán rất khác nhau. Công nghệ định vị bằng vệ tinh GNSS với các chùm vệ tinh nổi tiếng như GPS của Mỹ, GLONASS của Nga, ... đã có thể đạt đến độ chính xác 1cm với thuật toán RTK. Với độ chính xác cao như vậy, nó hoàn toàn có thể đáp ứng cho các bài toán liên quan đến điều khiển từ xa. Tuy nhiên GNSS chỉ thực sự mạnh khi ở không gian bên ngoài trời, còn khi ở trong nhà, GNSS có độ chính xác rất thấp, thậm chí là tín hiệu không đủ mạnh để có các thông tin về vị trí. Từ đó, các công nghệ không dây có thể hoạt động trong nhà và đã có sẵn như WiFi, Bluetooth, Bluetooth Low Energy được xét đến

Bài toán định vị trong nhà là một trong những vấn đề quan trọng trong lĩnh vực IoT và được ứng dụng rộng rãi trong các ứng dụng như giám sát và điều khiển các hệ thống tự

động trong nhà, hỗ trợ điều hướng trong các khu vực công cộng như sân bay, bệnh viện, trung tâm mua sắm, và cải thiện trải nghiệm người dùng.

Tuy nhiên, định vị trong nhà đối mặt với nhiều thách thức vì tín hiệu từ GPS không thể xuyên qua tường và được giảm thiểu đáng kể trong nhà. Vì vậy, các phương pháp định vị trong nhà thường sử dụng các công nghệ không dây như Wi-Fi, Bluetooth, Zigbee, RFID, Ultra-Wideband (UWB) để xác định vị trí của thiết bị hoặc người dùng trong một khu vực bao quanh.

Trong đó, công nghệ định vị bằng Wi-Fi và Bluetooth được sử dụng phổ biến nhất trong các ứng dụng định vị trong nhà. Khi sử dụng Wi-Fi, thiết bị sẽ đọc tín hiệu Wi-Fi Access Point (AP) trong khu vực và dựa trên đó để xác định vị trí của thiết bị hoặc người dùng. Còn đối với Bluetooth, các thiết bị sẽ truyền dữ liệu định vị giữa chúng để xác định khoảng cách và vị trí của thiết bị hoặc người dùng.

Các phương pháp định vị trong nhà bằng Wi-Fi và Bluetooth thường sử dụng các thuật toán như trilateration và fingerprinting để xác định vị trí của thiết bị hoặc người dùng. Tuy nhiên, các phương pháp này vẫn đối mặt với nhiều thách thức như ảnh hưởng của tường, nhiễu sóng và độ chính xác không đáp ứng được yêu cầu của một số ứng dụng đòi hỏi độ chính xác cao.

1.2 Tổng quan về công nghệ WiFi và Bluetooth

1.2.1 Tổng quan về công nghệ định vị

Công nghệ WiFi và Bluetooth đều là các công nghệ không dây được sử dụng rộng rãi trong các thiết bị di động, máy tính và các thiết bị IoT. Cả hai công nghệ đều hoạt động trong băng tần ISM 2,4 GHz và được thiết kế để cho phép truyền dữ liệu giữa các thiết bị. Công nghệ WiFi là một phương tiện truyền thông không dây được sử dụng rộng rãi để kết nối internet và truyền dữ liệu giữa các thiết bị trong mạng LAN (Local Area Network). Để định vị vị trí trong nhà bằng WiFi, các thiết bị sử dụng thông tin về tín hiệu WiFi được phát ra từ các điểm truy cập (access points) trong mạng. Các thông tin này bao gồm tên điểm truy cập, độ mạnh tín hiệu, tần số và thời gian phản hồi. Thông qua việc phân tích và xử lý

các thông tin này, các thiết bị định vị có thể tính toán được vị trí của chúng trong không gian. Công nghệ Bluetooth là một công nghệ truyền thông không dây cho phép truyền dữ liệu giữa các thiết bị trong khoảng cách gần nhau. Trong định vị vị trí trong nhà bằng Bluetooth, các thiết bị sử dụng thông tin về tín hiệu Bluetooth được phát ra từ các thiết bị khác để xác định vị trí của chúng. Các thông tin này bao gồm tên thiết bị, độ mạnh tín hiệu và tần số. Các thiết bị định vị có thể tính toán được vị trí của chúng trong không gian dựa trên thông tin về độ mạnh tín hiệu Bluetooth và khoảng cách giữa các thiết bị.

Công nghệ WiFi và Bluetooth là hai công nghệ được sử dụng phổ biến trong định vị trong nhà (indoor positioning). Cả hai công nghệ này sử dụng tín hiệu sóng vô tuyến để xác định vị trí của thiết bị trong một không gian bao gồm nhiều điểm truy cập (access point) hoặc thiết bị phát sóng (beacon) được đặt ở các vị trí khác nhau trong không gian. Trong công nghệ định vị bằng WiFi, tín hiệu sóng WiFi được sử dụng để xác định vị trí của thiết bị. Khi thiết bị di chuyển trong không gian, tín hiệu sóng WiFi sẽ tương tác với các điểm truy cập WiFi (access point) được đặt trong phạm vi định vị, từ đó thu thập thông tin về chất lượng tín hiệu (RSSI) và độ trễ (latency) để tính toán vị trí của thiết bị. Công nghệ định vị bằng WiFi được sử dụng phổ biến trong các ứng dụng như định vị nhân viên trong công ty, định vị sản phẩm trong siêu thị, định vị trong bệnh viện, vv. Trong công nghệ định vị bằng Bluetooth, tín hiệu sóng Bluetooth được sử dụng để xác định vị trí của thiết bị. Thiết bị phát sóng Bluetooth (beacon) được đặt ở các vị trí khác nhau trong không gian, từ đó thiết bị nhận được tín hiệu sóng Bluetooth để tính toán vị trí của thiết bị. Công nghệ định vị bằng Bluetooth được sử dụng phổ biến trong các ứng dụng như định vị trong bãi đậu xe, định vị trong các sự kiện lớn, vv.

1.2.2 Phương pháp công nghệ định vị

1.2.2.1 Time and space attributes of received signal (TSARS)

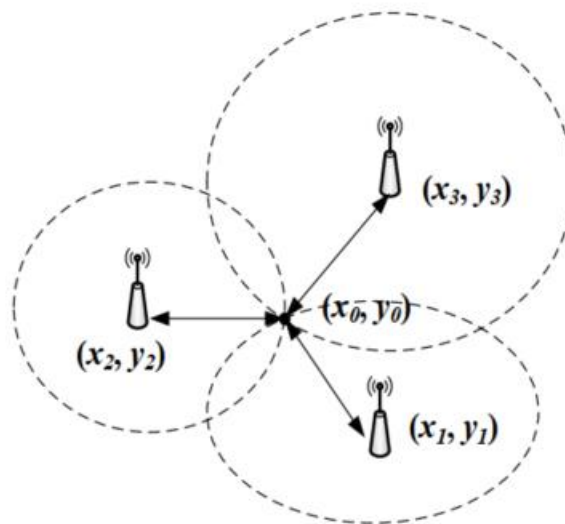
Phương pháp TSARS bao gồm các thuật toán như sau:

- *ToA(Timer of Arrival)*: tính toán khoảng cách dựa vào thời gian đến của tín hiệu được gửi bởi AP

- *TDoA*(Time Difference Of Arrival): dựa trên việc tính toán chênh lệch thời gian đến của các AP.
- *AoA*(Angle of Arrival): đo góc của tín hiệu đến từ đó tính được khoảng cách của thiết bị

1.2.2.2 Received-signal strength (RSS)

Trilateration: là thuật toán tính toán vị trí thông qua việc biến đổi cường độ tín hiệu nhận được bởi người dùng (RSS) thành khoảng cách trong không gian và coi khoảng cách đó là bán kính của 1 hình tròn, sau đó lấy giao điểm của các đường tròn đó là vị trí người dùng. Thuật toán Trilateration yêu cầu từ 3 AP và có thể kết hợp giữa KNN và thuật toán Fingerprint có thể cho độ chính xác lên tới 0.5m.

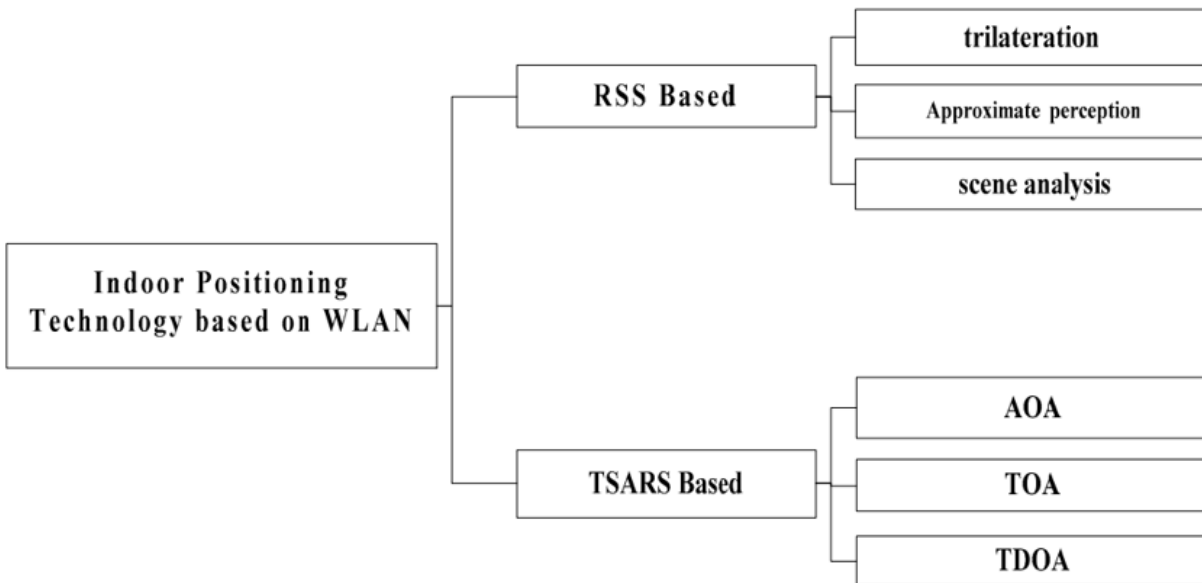


Hình 1. 1 Phương pháp Triangulation

Approximate Perception: sử dụng AP có cường độ mạnh nhất làm tiêu chí định vị do đó cần phải biết vị trí chính xác của AP, độ chính xác thường không được cao.

Scene Analysis: được coi là thuật toán Fingerprint và không cần biết vị trí của các AP. Thuật toán này dựa trên cường độ tín hiệu các điểm AP khác nhau tại cùng 1 điểm RP được đo và thu được thông tin cường độ của các RP. Kết hợp với thông tin tọa độ và địa chỉ MAC (Media Access Control) của RP, sau đó thành lập 1 Fingerprint database hay còn

được gọi là radio map. Với Radio Map có thể so sánh cường độ tín hiệu người dùng với database và sử dụng các thuật toán để thu được vị trí người dùng. Thuật toán Scene analysis chia làm 2 pha đó là pha Offline - thiết lập Fingerprint database; pha Online - sử dụng các thuật toán để tìm ra vị trí.



Hình 1. 2 Cấu trúc của công nghệ định vị bằng WiFi

Thuật toán fingerprint dựa trên công nghệ định vị bằng wifi là 1 trong các công nghệ định vị phổ biến nhất dựa trên model 2D, và được sử dụng cả trong định vị ngoài trời và trong nhà.

- Đối với định vị ngoài trời thuật toán fingerprint được chia ra làm 3 loại: visual fingerprint sử dụng hình ảnh; motion fingerprint sử dụng cảm biến chuyển động như cảm biến gia tốc, la bàn điện tử; signal fingerprint sử dụng RSSI(Received Signal strength indication).
- Đối với định vị trong nhà chủ yếu sử dụng signal fingerprint. Công nghệ sử dụng cường độ tín hiệu nhận được của mỗi vị trí (được lưu trong database) để đối chiếu với cường độ tín hiệu đo được tại vị trí users để đưa ra vị trí chính xác.

CHƯƠNG 2: ĐỀ XUẤT CÁC THUẬT TOÁN, PHƯƠNG PHÁP

2.1 Đánh giá một vài phương pháp định vị

Trong các mô hình định vị trong nhà với không gian vừa, nhiều phương pháp định vị đã được đề xuất và thử nghiệm.

Bài báo [1] sử dụng các Access Point (AP) cố định là các bộ thu và thiết bị cầm tay là bộ phát. Bằng cách sử dụng việc thu tín hiệu RSSI và áp dụng phương pháp SDOR để lọc tín hiệu RSSI, thuật toán Least Square và Centroid Positioning để tính toán ra các tọa độ trên mặt phẳng 2D. Theo kết quả thực nghiệm được ghi, độ chính xác có thể đạt được từ 1m đến 1.5m. Bài báo [2] áp dụng phương pháp tương tự bài báo [1] và cũng đạt được độ chính xác khoảng 1m. Việc định vị không chỉ riêng ở xác định RSSI của tín hiệu cao tần, mà có thể sử dụng cả tín hiệu âm thanh như bài báo [3]. Trong bài báo này, việc kết hợp giữa tín hiệu Bluetooth Low Energy và sóng âm được áp dụng. Thiết bị để thực hiện thực nghiệm có khả năng phát cả sóng âm và tín hiệu Bluetooth Low Energy để có thể chuyển đổi giữa 2 chế độ, phòng khi 1 trong 2 bị mất tín hiệu. Các bài báo đã được nêu trên đều là việc thực hiện định vị trực tuyến (on-line) đơn thuần. Tuy nhiên, định vị có thể đạt được độ chính xác hơn, khi thực hiện một vài thao tác ngoại tuyến (off-line) như việc áp dụng phương pháp dấu vân tay. Bài báo [4] áp dụng dấu vân tay kết hợp với thuật toán K-nearest Neighbors (KNN) để đạt được độ chính xác dưới 1m. Các Access Point sử dụng như các điểm phát, được đặt cố định. Đồng thời, không gian được mô hình hóa lưới điểm với các điểm cách nhau 1m. Việc thu tín hiệu trong giai đoạn ngoại tuyến (off-line) được thực hiện với từng điểm lưới để vẽ ra bản đồ véc-tơ cho từng điểm. Sau khi xong giai đoạn ngoại tuyến, giai đoạn trực tuyến được thực hiện sử dụng phương pháp K-nearest Neighbors (KNN). Bài báo [5] cũng áp dụng phương pháp tương tự, nhưng số lượng Access Point bị giảm đi và vùng không gian thực nghiệm cũng lớn hơn nhiều, gây ra sai số lớn hơn ở bài báo trước là 2.6m khi diện tích chứa 1 Access Point là 30m² và 4.8m khi diện tích đó là 100m². Phương pháp tương tự cũng được áp dụng trong bài báo [6], tuy nhiên,

việc sử dụng Wifi để định vị được xem là chính, còn Bluetooth Low Energy chỉ được sử dụng khi không có Wifi trong vùng.

Bài báo [7] [8] dùng phương pháp dấu vân tay, đồng thời bổ sung thêm các thuật toán để cập nhật lại dữ liệu được đo ở giai đoạn ngoại tuyến. Phương pháp Polynomial Regresstion Model (PRM) được áp dụng trong [7] và phương pháp Extended Kalman Filtering (EKF) được áp dụng trong cả 2 bài. Trong bài [8], phương pháp Location of Interest (LoI) được áp dụng để tránh việc sai lệch quá lớn.

Cơ sở dữ liệu mẫu [9] được áp dụng để tính toán sai số của các phương pháp theo lý thuyết và đạt được các kết quả với việc áp dụng số lượng điểm tham chiếu là 3 như sau:

Bảng 1. 1 Kết quả tính toán sai số theo lý thuyết

Thuật toán		Sai số nhỏ nhất	Sai số lớn nhất	Sai số trung bình
Giao khoảng cách	LS	0.31152	7.36744	3.73543
	BGI (3 điểm)	0.34282	4.40103	2.21919
Fingerprint	KNN (3 điểm)	0.46149	3.26193	1.65332
	KNN + BGI (3 điểm)	0.61163	3.68050	1.83927
	WKNN (3 điểm)	0.47749	2.91185	1.57001

Với trường hợp chỉ sử dụng 3 điểm đánh giá thì trong các thuật toán, sai số của phương pháp WKNN là bé nhất với cơ sở dữ liệu được lấy từ dataset. Tuy nhiên vẫn chưa đánh giá được tất cả các trường hợp với số lượng điểm sử dụng để xác định vị trí là lớn hơn 3.

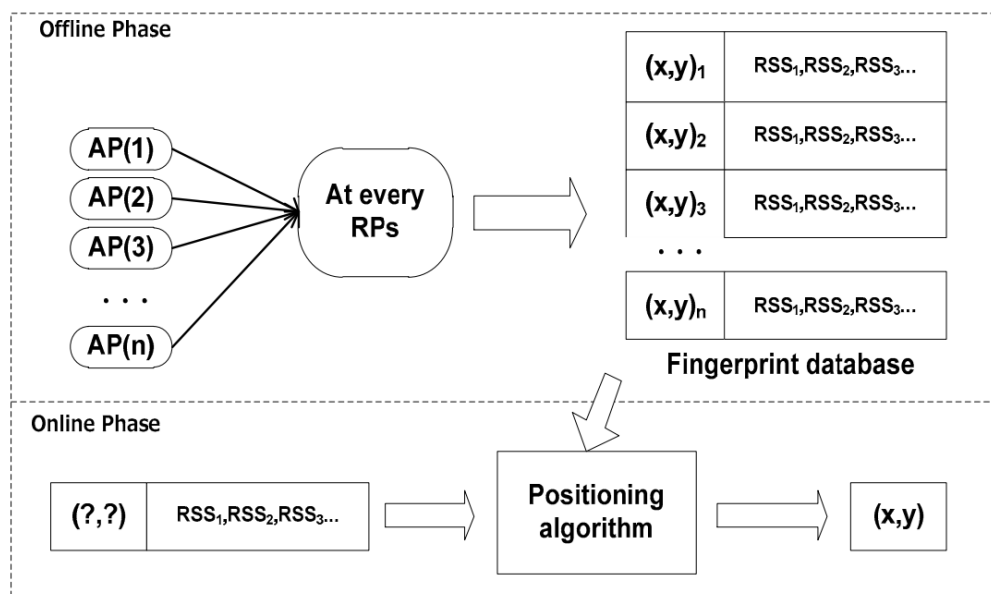
Trong bài báo [10], Bai et al đã sử dụng phương pháp kết hợp thuật toán KNN và phương pháp Fingerprint cùng với thuật toán Triangular có hiệu quả sai số giảm 0.5m so với hệ thống phương pháp định vị thời gian thực của Ekahau[11].

2.2 Phương pháp định vị Fingerprint

2.2.1 Tổng quan về phương pháp Fingerprint

Phương pháp fingerprint là một phương pháp định vị dựa trên các vân tay tín hiệu (signal fingerprints) của các bộ phát sóng wifi (access points) được phân bố trong khu vực cần định vị. Các vân tay tín hiệu này là một tập hợp các thông tin định danh của các access points và các thông số tín hiệu như độ mạnh (RSSI), độ trễ (latency), độ rộng (bandwidth), ... được thu thập từ các điểm dữ liệu trên toàn bộ khu vực cần định vị.

Các data Fingerprint được lưu trữ trong một cơ sở dữ liệu và sử dụng để so khớp với các vân tay tín hiệu mới được thu thập từ điện thoại hoặc thiết bị định vị. Bằng cách so khớp các thông tin định danh và tín hiệu của các access points trong vân tay tín hiệu mới với các vân tay tín hiệu trong cơ sở dữ liệu, phương pháp fingerprint cho phép xác định vị trí của thiết bị với độ chính xác tương đối cao.



Hình 2. 1 Mô hình tổng quan của phương pháp Fingerprint

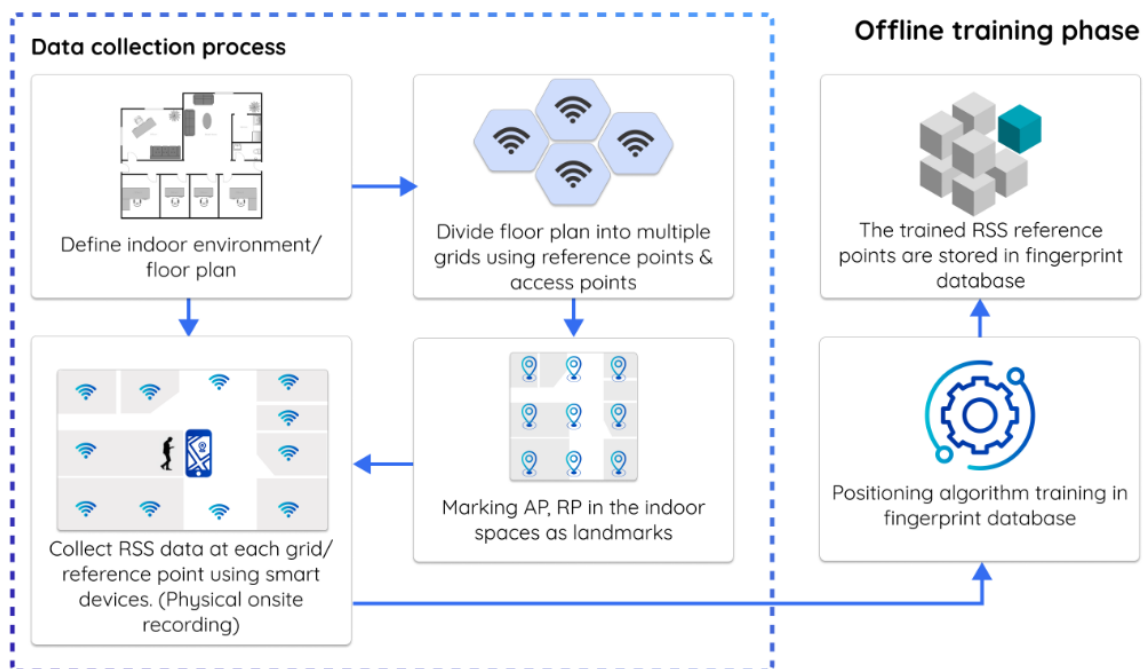
Nguyên lý hoạt động của một hệ thống sử dụng phương pháp Fingerprint gồm có 2 pha:

Offline phase trong fingerprint-based indoor localization là giai đoạn tiền xử lý dữ liệu, trong đó các dữ liệu thu thập được từ các cảm biến được xử lý và phân tích để tạo ra bản đồ vân tay (fingerprint map) của môi trường. Trong giai đoạn này, người dùng thực hiện việc thu thập dữ liệu từ các cảm biến (chẳng hạn như các điểm truy cập Wi-Fi hoặc Bluetooth) trong khu vực cần định vị. Dữ liệu được thu thập bao gồm tín hiệu thu được từ các cảm biến và vị trí tương ứng của người thu thập. Sau đó, dữ liệu được xử lý để tạo ra bản đồ vân tay, cung cấp thông tin về tín hiệu từ các cảm biến tại từng vị trí trong môi trường.

Các bước chính trong giai đoạn offline bao gồm:

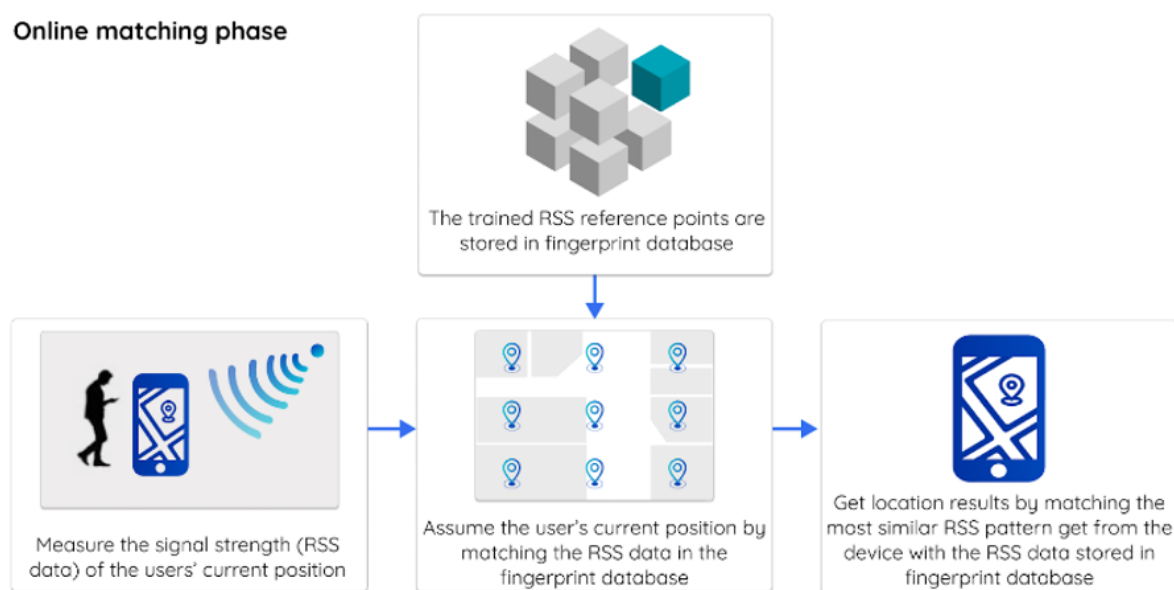
- Thu thập dữ liệu từ các cảm biến trong khu vực định vị.
- Xử lý và lưu trữ dữ liệu để tạo ra bản đồ vân tay.
- Xác định các điểm đặt cảm biến.

Bản đồ vân tay được tạo ra trong giai đoạn offline sau đó được sử dụng trong giai đoạn online để định vị vị trí của người dùng.



Hình 2. 2 Mô hình thuật toán sử dụng trong pha Offline

Online phase là giai đoạn đưa ra quyết định vị trí của người dùng dựa trên dữ liệu thu thập được từ tín hiệu wifi tại thời điểm hiện tại. Trong giai đoạn này, một bộ phân loại sẽ được sử dụng để phân loại vị trí của người dùng bằng cách so sánh dữ liệu thu thập được với các dữ liệu được lưu trữ trong cơ sở dữ liệu từ giai đoạn offline. Trong pha online, hệ thống sẽ tiếp nhận dữ liệu wifi thu thập được từ thiết bị di động của người dùng tại thời điểm hiện tại, sau đó tiến xử lý và trích xuất các đặc trưng từ dữ liệu này. Sau đó, các đặc trưng này được đưa vào bộ phân loại để phân loại vị trí của người dùng. Các kết quả định vị được trả về cho người dùng thông qua giao diện người dùng hoặc tích hợp vào các ứng dụng. Pha này thường được thực hiện bằng cách sử dụng các thuật toán như KNN (K-Nearest Neighbors), SVM (Support Vector Machine), Random Forest, Neural Networks và các phương pháp tối ưu hóa khác. Kết quả định vị sẽ được hiển thị trên bản đồ hoặc trả về cho người dùng.



Hình 2. 3 Mô hình trong pha Online

2.2.2 Các nhân tố ảnh hưởng tới phương pháp định vị Fingerprint

Effects of the Human Body on Signal Strength: do cơ thể con người 70% là nước, mà nước có thể hấp thụ tín hiệu sóng RF ở tần số 2.4Ghz gây ra sự sụt giảm và nhiễu tín hiệu.

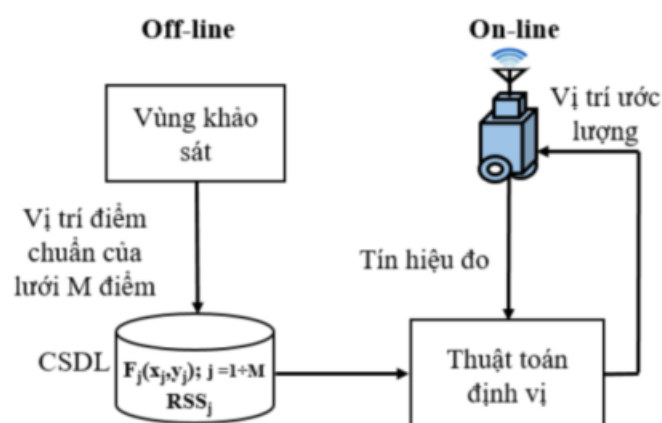
Effects of the Multipath Effect on Signal Strength: xảy ra khi không gian trong nhà có nhiều vật cản làm cho tín hiệu sóng RF lan truyền qua các đường khác nhau gây ra việc tín hiệu đến thiết bị nhận tại những thời điểm khác nhau, dẫn đến lệch pha, chồng chéo và sai lệch tín hiệu. Multipath effect gồm có 2 loại the convolution filter effect; communication noise. Người ta lập mô hình cả 2 loại với mô hình lan truyền logarit làm giảm xác suất sai lệch chuẩn của việc estimate từ 42% xuống 29%

Effects of the Number of APs and RPs: số lượng của AP và RP có ảnh hưởng tới độ chính xác của dữ liệu, thời gian xử lý trong pha Offline và giá cả. Theo nghiên cứu trong phạm vi khoảng 11x23m thì số lượng AP từ 5 trở lên thì kết quả không thay đổi; số lượng RP từ 66 trở lên cũng không thay đổi.

Effects of Mobile Devices: do việc sản xuất các thiết bị nhận là khác nhau nên có các chuẩn khác nhau để convert số liệu đến đến việc đo cùng 1 vị trí nhưng với thiết bị khác nhau là khác nhau.

2.3 Đề xuất các thuật toán và phương pháp định vị

Sử dụng kỹ thuật quét 3 kênh quảng bá của WiFi được sử dụng trong để giảm đi ảnh hưởng của hiệu ứng đa đường đối với kết quả thu được. Việc xây dựng bản đồ được tạo thành 3 bản ứng tương đương với 3 kênh quảng bá này.



Hình 2. 4 Hệ thống định vị theo phương pháp Fingerprint

2.3.1 Thuật toán KNN truyền thống

Thuật toán KNN giúp phân loại mẫu dựa trên phép đo, tính toán các mẫu có sẵn. Tại đây, tập khoảng cách Euclidean D_i giữa phép đo các RSS $S = \{s_{i1}, s_{i2}, \dots, s_{im}\}$ và phép đo các RSS trong lưới các dấu vân tay tham chiếu $R_i = \{r_{i1}, r_{i2}, \dots, r_{im}\}$ sẽ được tính toán. Tập các khoảng cách D_i được sắp xếp để tìm ra K dựa vào những khoảng D_i nhỏ nhất. Toạ độ của đối tượng được tính toán dựa vào toạ độ trung bình của K dấu vân tay (giai đoạn online). Ta có công thức tính khoảng cách Euclidean được tính như sau:

$$D_i = \sqrt{\sum_{j=1}^N |RSS_j - RSS_{ij}|^2}, \text{ với } i = 1, 2, \dots, M \quad (1)$$

Ở đây, RSS_{ij} đại diện cho giá trị trung bình của cường độ tín hiệu nhận trên điểm thử nghiệm thứ j được tính toán với điểm tham chiếu thứ i trên bản đồ dấu vân tay tham chiếu và RSS_j đại diện cho giá trị RSS của AP thứ j thu được trong giai đoạn thử nghiệm trực tuyến. M và N biểu thị tương ứng cho các điểm tham chiếu (RPs) và các điểm thử nghiệm. Tiếp theo đó, chọn ra k kết quả nhỏ nhất trong tập D_i ta được k toạ độ của các điểm tham chiếu (RP). Từ đó xác định được toạ độ vị trí của đối tượng bằng công thức:

$$(x, y) = \frac{1}{k} \sum_{i=1}^k (x_i, y_i) \quad (2)$$

Trong đó (x, y) là toạ độ điểm cần tìm, k có thể được ước lượng

Tài liệu [4] sử dụng các Access-Point (AP) cố định đặt ở các vị trí bao toàn phòng. Trong giai đoạn off-line sẽ lấy giá trị của lưới điểm cách nhau 1m để tạo thành các vecto RSSI cho mỗi điểm lưới đó và tạo thành một hệ bản đồ tín hiệu. Mỗi vecto tại một điểm bao gồm nhiều giá trị RSSI tương đương với từng Access-Point được sử dụng. Sai số của phương pháp phụ thuộc vào độ dày đặc của lưới điểm và thường nhỏ hơn so với khoảng cách của 2 điểm cạnh nhau trong lưới.

2.3.2 Thuật toán WKNN

Giai đoạn ngoại tuyến (off-line):

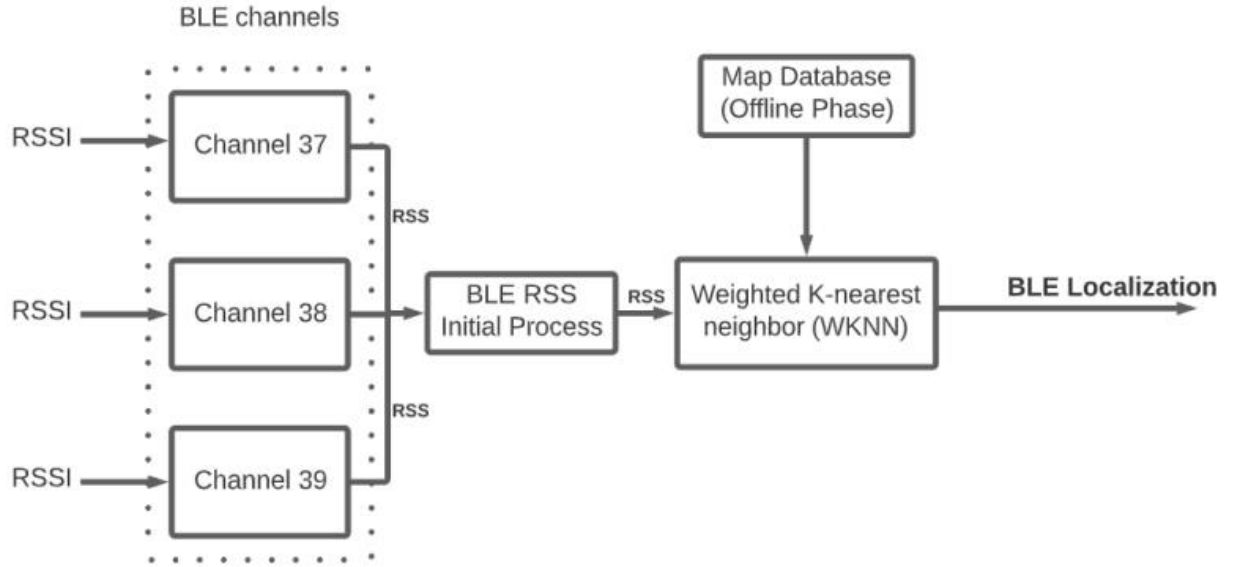
Tại mỗi điểm tham chiếu trong phạm vi khu vực định vị, chỉ số cường độ tín hiệu nhận được (RSSI) của tất cả các trạm phát đối với điểm đó sẽ được thu thập lại và một cơ sở dữ liệu dấu vân tay được hình thành dựa trên tất cả các chỉ số thu được bao gồm tất cả các tọa độ vị trí của các điểm tham chiếu (hay dấu vân tay tham chiếu). Giả sử $A_{ij}[\varepsilon]$ là giá trị RSS của điểm tham chiếu thứ j được thu thập từ trạm phát thứ i ở thời điểm ε . Dữ liệu của các dấu vân tay được thiết lập trong giai đoạn ngoại tuyến được biểu thị bởi:

$$A = \begin{pmatrix} A_{1,1}[\varepsilon] & A_{2,1}[\varepsilon] & \cdots & A_{1,N}[\varepsilon] \\ A_{2,1}[\varepsilon] & A_{2,2}[\varepsilon] & \cdots & A_{2,N}[\varepsilon] \\ \vdots & \vdots & \ddots & \vdots \\ A_{M,1}[\varepsilon] & A_{M,2}[\varepsilon] & \cdots & A_{M,N}[\varepsilon] \end{pmatrix} \quad (3)$$

Trong đó M là số trạm phát, N là số điểm tham chiếu (RP), $\varepsilon = 1, 2, \dots, E$ là thời điểm thiết lập chỉ số RSS mẫu, E là tổng số thời điểm mà điểm tham chiếu thứ j được thu thập bởi trạm phát thứ i ở những thời điểm khác nhau. Cột vector của A đại diện cho tất cả các giá trị RSS của những trạm phát tại RP thứ j là toàn bộ ma trận của điểm tham chiếu, có thể được biểu thị bởi:

$$A_j = [A_{1,j}, A_{2,j}, \dots, A_{M,j}]^T \quad (4)$$

Giai đoạn trực tuyến (online)



Hình 2. 5 Phương pháp định vị WKNN theo BLE

Tại đây, thiết bị di động gửi giá trị RSS được đo từ M trạm phát với thiết bị thu và khớp dữ liệu với thuật toán để có được tọa độ thực tế của đối tượng. Để cải thiện độ chính xác về tọa độ của đối tượng, ta tính toán tất cả số liệu giữa ma trận chỉ số RSS nhận được trong giai đoạn ngoại tuyến và ma trận chỉ số RSS nhận được trong giai đoạn trực tuyến, sau đó chọn ra k RP lân cận với điểm thử nghiệm trong giai đoạn trực tuyến từ đó xác định hệ số gia trọng W (đây chính là sự khác biệt với thuật toán KNN truyền thống). Các bước cụ thể như sau. Giá trị RSS của l lần đo ở thời gian thực (thử nghiệm) được biểu diễn bởi:

$$\Phi_l = \begin{pmatrix} \phi_{1,l}[1] & \phi_{1,l}[2] & \cdots & \phi_{1,l}[E] \\ \phi_{2,l}[1] & \phi_{2,l}[2] & \cdots & \phi_{2,l}[E] \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{M,l}[1] & \phi_{M,l}[2] & \cdots & \phi_{M,l}[E] \end{pmatrix}_{M \times E} \quad (5)$$

Với $\phi_{M,l}$, $[E]$ là giá trị RSS đo được trong giai đoạn trực tuyến từ M trạm phát được thu thập ở E thời điểm. Từ đó ta tính giá trị RSS trung bình của trạm phát thứ i được nhận bởi điểm thử thứ l ở bất kì thời điểm nào, được biểu diễn bởi vector:

$$\overline{\Phi_l} = [\overline{\phi_{1,l}} \quad \overline{\phi_{2,l}} \quad \cdots \quad \overline{\phi_{i,l}} \quad \cdots \quad \overline{\phi_{M,l}}]^T \quad (6)$$

$$\overline{\phi_{i,l}} = \frac{1}{T} (\phi_{i,l}[1] + \phi_{i,l}[2] + \dots + \phi_{i,l}[E]) \quad (7)$$

Kết hợp với CSDL dấu vân tay được thu thập ở giai đoạn ngoại tuyến sẽ xây dựng được ma trận khoảng cách của RSS như sau:

$$S = \begin{pmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,N} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ s_{M,1} & s_{M,2} & \cdots & s_{M,N} \end{pmatrix} \quad (8)$$

$$\hat{S} = [\hat{s}_1 \quad \hat{s}_2 \quad \dots \quad \hat{s}_j \quad \dots \quad \hat{s}_N] \quad (9)$$

Từ tập \hat{S} , sắp xếp từ nhỏ đến lớn và chọn K giá trị nhỏ nhất. Tập \hat{s}^j được biểu thị bởi r_1, r_2, r_3, \dots , và tọa độ các điểm tham chiếu tương ứng là (x_k, y_k) , $k=1, 2, \dots, K$. Thuật toán KNN dựa vào cường độ tín hiệu nhận (RSS) và dùng giá trị trung bình để tính toán. Tuy nhiên, phân phối của RSS không phải lúc nào cũng là phân phối chuẩn (Gaussian) đối với môi trường trong nhà phức tạp có nhiều vật cản. Để giải quyết vấn đề này, thuật toán Weighted KNN (WKNN) được đưa ra. Giả sử điểm thử nghiệm có tọa độ L:

$$L = (\hat{x}, \hat{y}) = \sum_{k=1}^K \omega_k * (x_k, y_k) \quad (10)$$

ω_k là hệ số gia trọng của từng điểm tham chiếu trong số K điểm gần nhất của điểm thử nghiệm, dùng để xác định tọa độ của điểm cần tìm, được tính như sau:

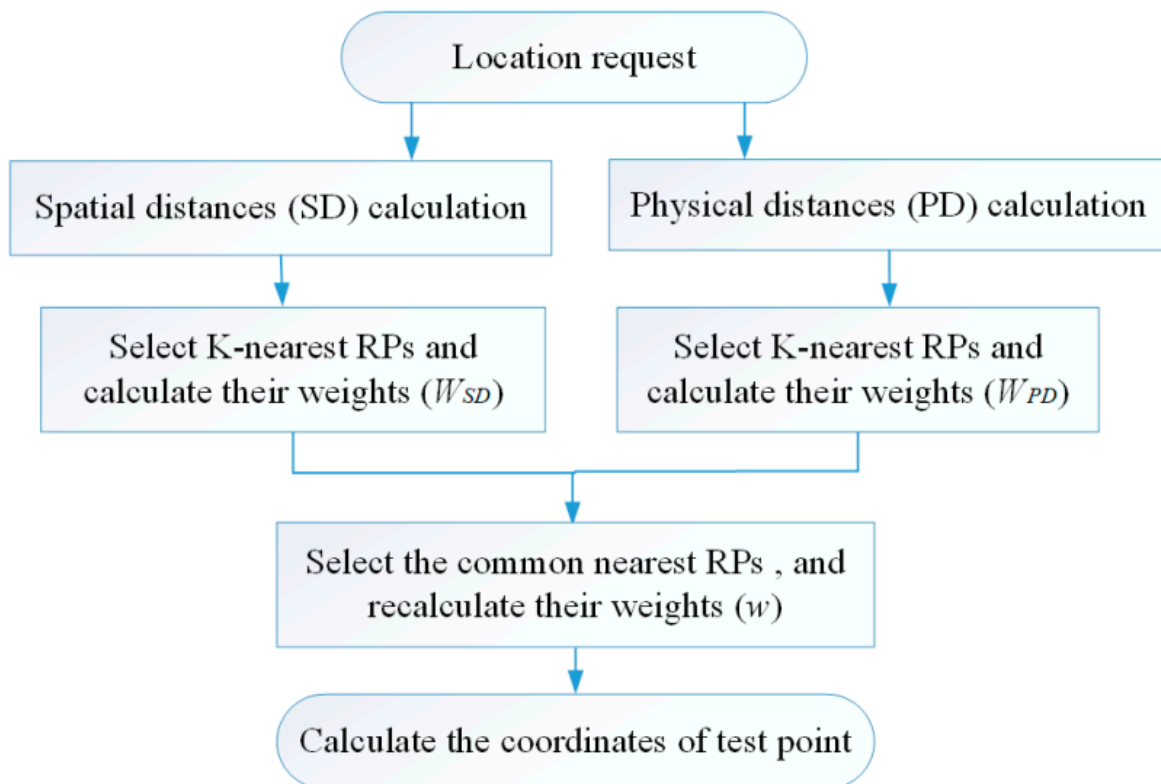
$$\omega_k = \frac{\frac{1}{r_k}}{\sum_{k=1}^K \left(\frac{1}{r_k}\right)} \quad (11)$$

Cuối cùng, vị trí của điểm thử nghiệm được đưa ra bởi công thức:

$$L = (\hat{x}, \hat{y}) = \sum_{k=1}^K \frac{\frac{1}{r_k}}{\sum_{k=1}^K \left(\frac{1}{r_k}\right)} * (x_k, y_k) \quad (12)$$

2.3.3 Thuật toán kết hợp giữa P-WKNN và S-WKNN

Việc kết hợp giữa 2 thuật toán P-WKNN (Physical weight K-nearest neighbor) và thuật toán S-WKNN (Spatial weight K-nearest neighbor) được đề xuất để cải thiện độ chính xác của vị trí. Ở đó việc tính toán khoảng cách không chỉ là tính khoảng cách Euclid trong không gian SD (Spatial distance) mà còn tính khoảng cách về vật lý PD (Physical distance). Sau đó chọn ra các RPs chung giữa K-nearest của cả 2 thuật toán, và những điểm RPs chung này được sử dụng để tính toán tọa độ của điểm TP.

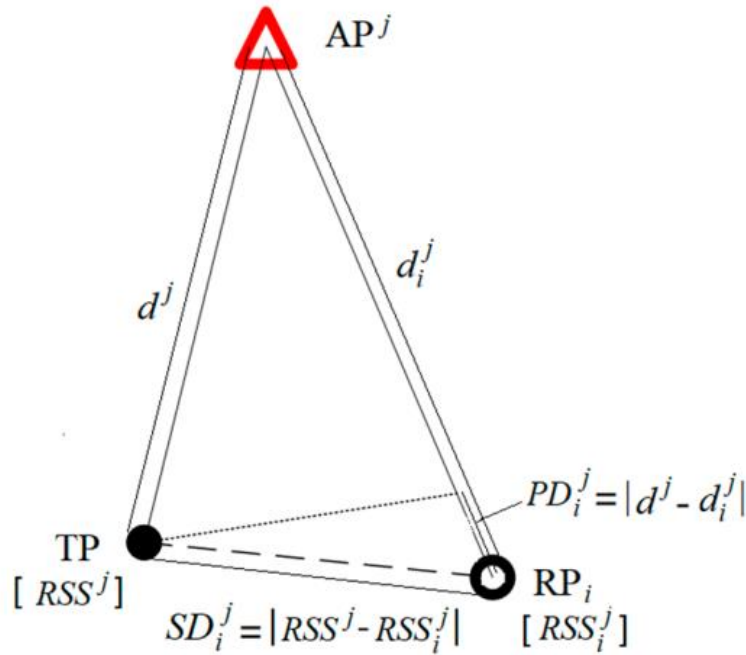


Hình 2. 6 Framework của thuật toán

2.3.3.1 Tính khoảng cách

Khoảng cách PD của RSSI được định nghĩa là trung bình cộng các khoảng cách từ TP tới các điểm AP và từ các điểm RP tới các điểm AP tương ứng.

Hình 2.7 biểu diễn việc tính khoảng cách vật lí và khoảng cách không gian. Hai khoảng cách này được tính dựa vào giá trị RSS. Sự khác nhau giữa 2 khoảng cách này là phương pháp tính toán. Khoảng cách không gian SD có mối tương quan tuyến tính với giá trị RSS trong khi khoảng cách vật lí có mối liên hệ là hàm mũ với giá trị RSS.



Hình 2. 7 Khoảng cách vật lí và khoảng cách không gian

Khoảng cách vật lí được tính theo công thức:

$$PD_i = \frac{\sum_{j=1}^m PD_i^j}{m} = \frac{\sum_{j=1}^m |d^j - d_i^j|}{m}, i = 1, 2, 3, \dots, N \quad (13)$$

Ở đó Pd_i là khoảng cách vật lí của RP thứ j , m là số lượng điểm AP mà có thể được phát hiện đồng thời tại TP và RP; giá trị của m có thể thay đổi giữa các điểm RP khác nhau; d_j là khoảng cách từ TP tới AP thứ j được tính bởi mô hình suy giảm tín hiệu.

$$d^j = d_0 10^{\left(\frac{RSS(d_0) - RSS^j}{10\eta}\right)} \quad (14)$$

Trong đó $RSS(d_0)$ là giá trị RSS đã đo tại khoảng cách d_0 từ AP thứ j ; RSS_j là giá trị RSS đã đo tại TP từ AP thứ j . d_j là khoảng cách từ RP thứ i tới AP thứ j được tính toán bởi mô hình suy giảm tín hiệu được tính theo công thức:

$$d_i^j = d_0 10^{\left(\frac{RSS(d_0) - RSS_i^j}{10\eta}\right)} \quad (15)$$

Trong đó RSS_{ij} là giá trị RSS đã đo tại RP thứ i từ AP thứ j .

Tất cả các khoảng cách vật lí được tính toán bởi công thức (13) và được lưu lại, sau đó chọn K điểm RP với khoảng cách PD nhỏ nhất. Trọng số của các điểm RP được chọn dựa vào khoảng cách vật lí được biểu diễn với $W(PD_i)$:

$$W_{PD_i} = \frac{\frac{1}{PD_i}}{\sum_{i=1}^K \frac{1}{PD_i}} \quad (16)$$

Còn khoảng cách không gian SP được tính theo công thức khoảng cách Euclide từ RP thứ i tới TP được biểu diễn bởi Sd_i :

$$SD_i = \sqrt{\frac{\sum_{j=1}^m (RSS^j - RSS_i^j)^2}{m}}, i = 1, 2, 3, \dots, N \quad (17)$$

Tất cả các khoảng cách RP được tính toán và lưu lại sau đó chọn ra K điểm RP với SD nhỏ nhất. Trọng số của các điểm RP được tính dựa trên khoảng cách không gian được biểu diễn bởi W(SD_i):

$$W_{SD_i} = \frac{\frac{1}{SD_i}}{\sum_{i=1}^K \frac{1}{SD_i}} \quad (18)$$

2.3.3.2 Tính trọng số và tọa độ

Trọng số của việc chọn \bar{K} chung được tính dựa trên W(PD) và W(SD), cả 2 trọng số này được theo công thức sau:

$$w_i = \frac{W_{PD_i} + W_{SD_i}}{\sum_{i=1}^{\bar{K}} (W_{PD_i} + W_{SD_i})} \quad (19)$$

$$w_i = \frac{W_{PD_i} W_{SD_i}}{\sum_{i=1}^{\bar{K}} (W_{PD_i} W_{SD_i})} \quad (20)$$

Từ các tính toán trên tọa độ của điểm TP được tính theo công thức:

$$(\hat{x}, \hat{y}) = \sum_{i=1}^{\bar{K}} w_i(x_i, y_i) \quad (21)$$

2.3.4 Các thuật toán nội suy

Việc sử dụng các thuật toán nội suy trong các bài toán có tập dữ liệu trong không gian không đầy đủ hoặc thiếu một số giá trị. Từ việc mở rộng các điểm dữ liệu giúp cho bao phủ được số điểm RP trong một không gian từ đó có thể cải thiện cũng như loại bỏ một số điểm có sai số lớn. Trong báo cáo việc thực hiện nội suy từ các điểm có thể chọn dựa vào người thiết kế về số lượng cũng như vị trí. Sau khi chọn được các điểm (x, y) sẽ thực hiện nội suy từ các điểm này ra RSSi tương ứng với số lượng AP. Từ các điểm đó sẽ được sử dụng để thêm vào với số điểm train ban đầu và tính toán tọa độ của TP.

2.3.4.1 Nội suy sử dụng thuật toán KNN

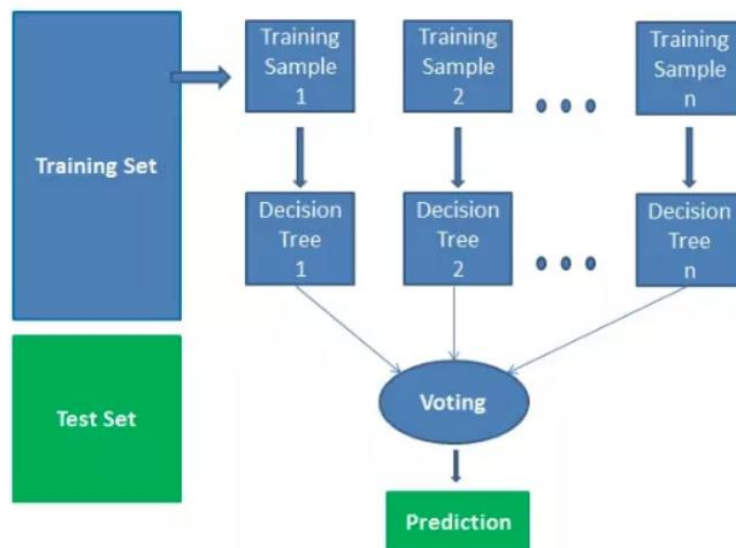
Thuật toán KNN về phương pháp và cơ sở đã được đề cập ở mục 2.3.1 và các công thức liên quan. Ý tưởng sử dụng thuật toán KNN vào việc nội suy các điểm RSSi tương ứng với AP thứ i từ các điểm (x, y)_i đã được chọn từ trước để mở rộng thêm số điểm train. Về quy tắc áp dụng sẽ ngược lại so với việc tìm ra vị trí (x, y) sử dụng RSS. Bây giờ đầu vào input của thuật toán là tọa độ đã định trước (x, y) sau đó output là giá trị cường độ tín hiệu RSS. Sau khi đã thực hiện nội suy, các điểm dữ liệu này được sử dụng kết hợp với thuật toán WKNN để tìm ra tọa độ của điểm TP.

2.3.4.2 Nội suy sử dụng thuật toán Random Forest

Random Forest là một thuật toán học máy dựa trên rừng quyết định (decision forest), được sử dụng rộng rãi trong nhiều ứng dụng khác nhau, bao gồm cả nội suy.

Random Forest là sự kết hợp của nhiều cây quyết định (decision trees) tạo thành một rừng quyết định (decision forest). Mỗi cây quyết định được xây dựng dựa trên một phần dữ liệu được lấy ngẫu nhiên từ tập dữ liệu huấn luyện. Quá trình huấn luyện của mỗi cây quyết định bao gồm các bước sau:

1. Chọn ngẫu nhiên một tập con của các tính năng để xây dựng cây.
2. Tìm ra thuộc tính (feature) tốt nhất để phân chia tập dữ liệu.
3. Phân chia tập dữ liệu thành hai tập con dựa trên thuộc tính đã chọn.
4. Lặp lại các bước trên cho các tập con dữ liệu mới cho đến khi đạt được điều kiện dừng (ví dụ như cây đã đạt đến độ sâu tối đa hoặc số lượng mẫu trong mỗi nút lá đã đạt đến ngưỡng).



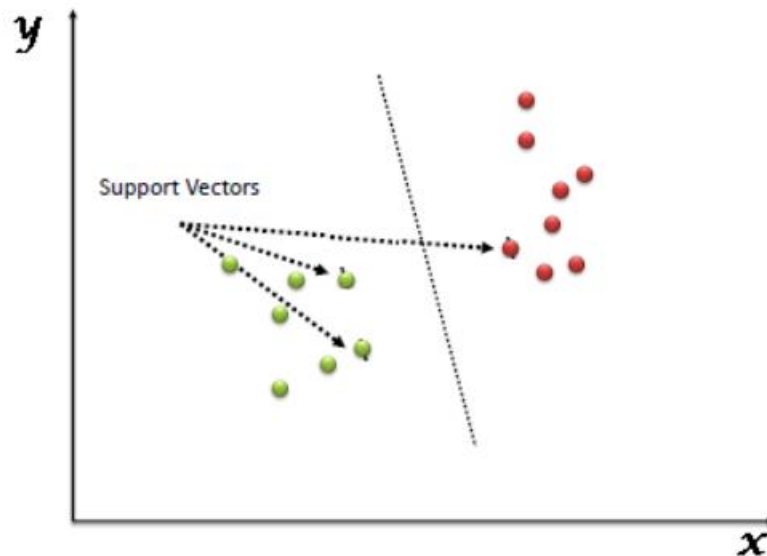
Hình 2. 8 Mô hình thuật toán Random forest

Khi sử dụng Random Forest để dự đoán giá trị của một điểm dữ liệu mới, thuật toán sẽ chạy dữ liệu đó qua mỗi cây quyết định trong rừng và tính toán giá trị dự đoán của từng cây. Cuối cùng, giá trị dự đoán của Random Forest sẽ là giá trị trung bình của các giá trị dự đoán của các cây quyết định. Để áp dụng Random Forest vào nội suy, ta có thể sử dụng các giá trị của các tính năng của các điểm xung quanh điểm cần được nội suy để huấn luyện một mô hình Random Forest. Sau khi huấn luyện xong, ta có thể sử dụng mô hình này để dự đoán giá trị của điểm cần được nội suy.

Một ưu điểm của Random Forest là nó có khả năng xử lý được các tính năng có mức độ tương quan cao và có thể sử dụng được với nhiều loại dữ liệu khác nhau.

2.3.4.3 Nội suy sử dụng thuật toán SVM

Trong định vị trong nhà, SVM (Support Vector Machine) cũng được sử dụng để thực hiện nội suy. SVM là một thuật toán học có giám sát và được sử dụng để phân loại và hồi quy dữ liệu. SVM có thể được áp dụng để tìm ra hàm xác định ranh giới phân chia giữa các lớp dữ liệu.



Hình 2. 9 Mô hình thuật toán SVM

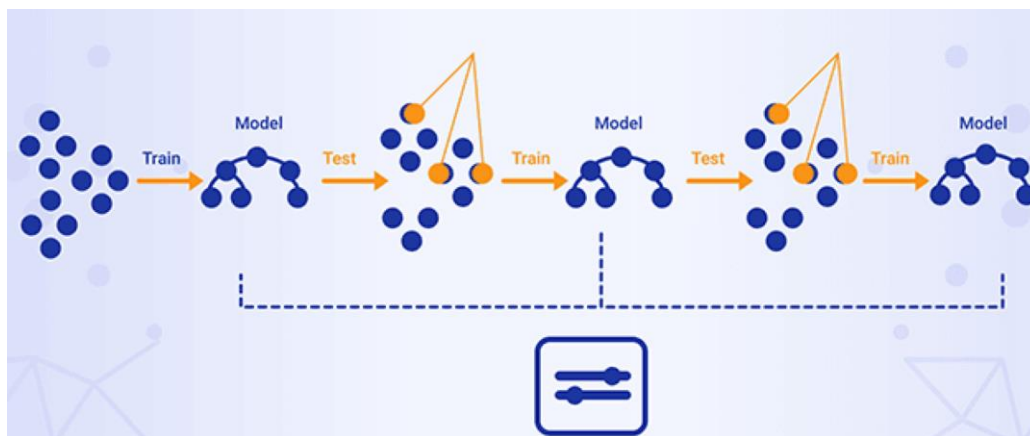
Các bước thực hiện nội suy sử dụng SVM như sau:

1. Chuẩn bị dữ liệu: Dữ liệu bao gồm các đặc trưng và vị trí của các điểm đo được. Dữ liệu được chia thành hai phần: tập huấn luyện và tập kiểm tra.
2. Huấn luyện SVM: Dữ liệu huấn luyện được sử dụng để tạo ra một mô hình SVM. Mô hình SVM được xây dựng để tìm ra một đường biên phân chia tốt nhất giữa các điểm dữ liệu.
3. Tính toán các trọng số: Sau khi mô hình SVM đã được huấn luyện, các trọng số của mô hình sẽ được tính toán. Các trọng số này được sử dụng để tính toán giá trị dự đoán của các điểm trong tập kiểm tra.
4. Tính toán vị trí: Sau khi các giá trị dự đoán đã được tính toán, thuật toán SVM được sử dụng để tìm ra vị trí của điểm cần được nội suy.

Thuật toán SVM được sử dụng rộng rãi trong các bài toán phân loại và hồi quy trong machine learning, và hiện nay cũng được sử dụng trong định vị trong nhà để thực hiện nội suy vị trí. Tuy nhiên, việc sử dụng SVM trong định vị trong nhà cần đảm bảo rằng các đặc trưng được chọn phù hợp và dữ liệu đầu vào phải được tiền xử lý tốt để đạt được độ chính xác cao.

2.3.4.4 Nội suy sử dụng thuật toán Gradient Boosting

Gradient boosting là một kỹ thuật học máy nhằm tạo ra một mô hình dự đoán tốt hơn bằng cách kết hợp nhiều mô hình dự đoán yếu. Thuật toán Gradient Boosting được sử dụng trong việc giải quyết bài toán hồi quy và phân loại.



Hình 2. 10 Mô hình thuật toán Gradient Boosting

Thuật toán Gradient Boosting có hai thành phần chính là Gradient descent và Boosting. Trong đó, Gradient descent là phương pháp tối ưu hóa mô hình bằng cách giảm thiểu lỗi giữa kết quả dự đoán và giá trị thực tế. Boosting là phương pháp kết hợp các mô hình dự đoán yếu để tạo ra một mô hình dự đoán mạnh hơn. Khi sử dụng Gradient Boosting để thực hiện nội suy, các bộ dữ liệu được sử dụng để huấn luyện mô hình dự đoán bao gồm tọa độ của các điểm đã biết và giá trị của các đại lượng cần nội suy tại các điểm đó. Sau khi huấn luyện, mô hình dự đoán sẽ được áp dụng để tính toán giá trị của đại lượng cần nội suy tại các điểm chưa biết.

Cụ thể, thuật toán Gradient Boosting sẽ thực hiện các bước sau:

1. Xác định số lượng các cây quyết định (decision tree) cần sử dụng.
2. Chia dữ liệu huấn luyện thành các tập con.
3. Đối với mỗi tập con, sử dụng Gradient descent để tối ưu hóa mô hình dự đoán.
4. Kết hợp các mô hình dự đoán yếu để tạo ra một mô hình dự đoán mạnh hơn.
5. Áp dụng mô hình dự đoán để tính toán giá trị của đại lượng cần nội suy tại các điểm chưa biết.

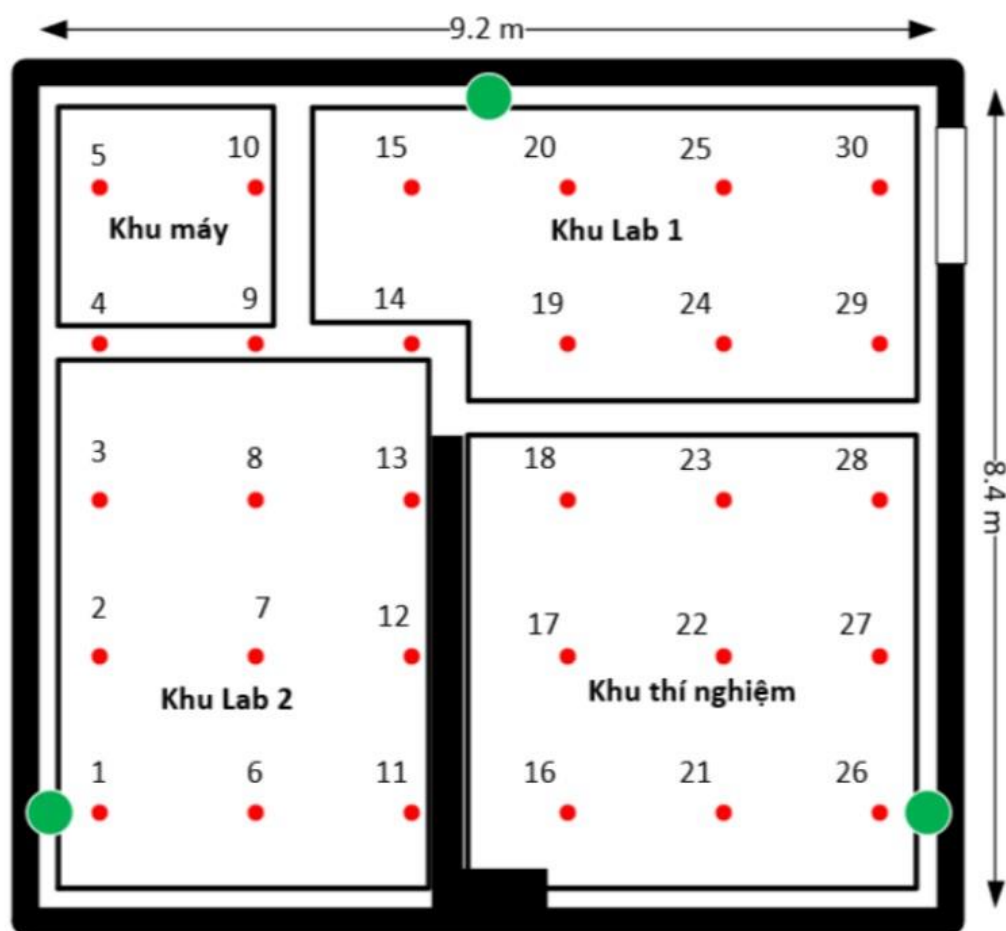
Tuy nhiên, việc sử dụng Gradient Boosting để thực hiện nội suy cần có một lượng dữ liệu đủ lớn và đa dạng để đảm bảo tính chính xác và độ tin cậy của kết quả nội suy.

CHƯƠNG 3: KẾT QUẢ THỬ NGHIỆM ĐỘ CHÍNH XÁC

3.1 Đánh giá sai số tập dữ liệu thứ nhất

Đối tượng thực nghiệm với tập dữ liệu thứ nhất. Địa điểm lựa chọn để đánh giá kết quả là phòng C1 – 328, trường Đại học Bách Khoa Hà Nội. Phòng được chia thành 4 khu vực: khu hoạt động Lab 1, khu hoạt động Lab 2, khu thí nghiệm Vi xử lý, khu máy chủ. Khoảng cách giữa các điểm lưới là 1.6m theo chiều dọc và chiều ngang. Tập dữ liệu thứ nhất gồm có 30 điểm train tương ứng là 30 RPs, 18 điểm test tương ứng là 18 TPps và 3 điểm phát tín hiệu RSS tương ứng là 3 Aps.

Việc thử nghiệm cần được đánh giá cùng với một vài phương pháp khác như KNN, WKNN, P&S-WKNN, các thuật toán nội suy.



Hình 3. 1 Đối tượng thực hiện thử nghiệm

Sai số các phương pháp:

Bảng 3. 1 Sai số tọa độ

Thuật toán	Sai số tọa độ (m)		
	Nhỏ nhất	Lớn nhất	Trung bình
KNN (30 điểm) k = 3	0.267	2.516	1.164
WKNN (30 điểm) k = 3	0.302	1.927	1.067
P&S-WKNN (30 điểm) k = 6	0.086	2.59	1.404
Gradient Boosting (99 điểm) + WKNN k = 4	0.295	1.828	1.389

Trong quá trình thực hiện thuật toán P&S-WKNN, các tham số của mô hình suy hao có giá trị như sau:

$$n = 4.92; P_{d0} = 34.4 \text{ dB}; d_0 = 1 \text{ m}$$

Sai số nội suy:

Bảng 3. 2 Sai số RSS

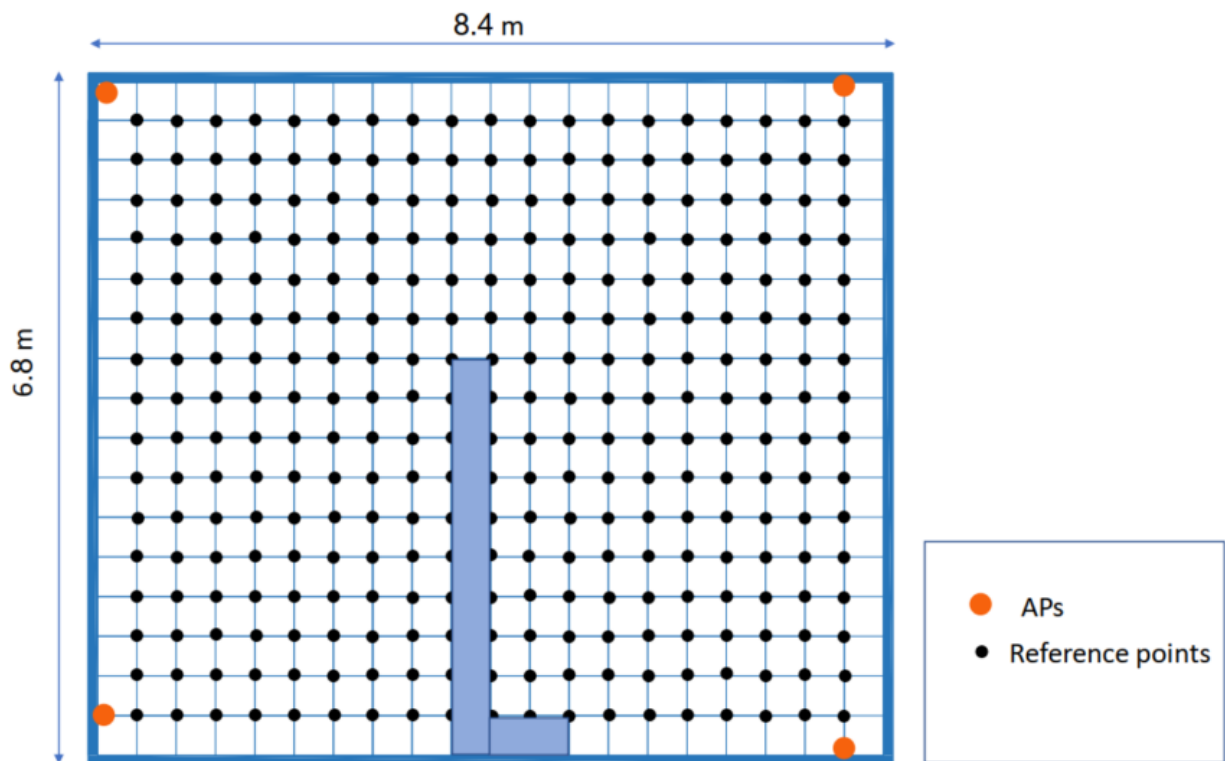
Thuật toán	Sai số RSS (dB)	
	Loss train	Loss valid
Gradient boosting	7.947	7.244
Random forest	6.102	9.372
SVM	11.326	8.811
KNN	9.584	9.264

Trong quá trình nội suy sử dụng thuật toán Gradient boosting, các siêu tham số được chọn có giá trị như sau:

$$\text{Loss} = 0.01; \text{learning_rate} = 1; n = 5;$$

3.2 Đánh giá sai số tập dữ liệu thứ hai

Đối tượng thực nghiệm của tập dữ liệu thứ hai giống với đối tượng thứ nhất là phòng C1 – 328, trường Đại học Bách Khoa Hà Nội. Nhưng số lượng điểm train và điểm test của tập dữ liệu thứ hai được đo lại với số lượng nhiều hơn so với số lượng dữ liệu của tập thứ nhất. Tập dữ liệu thứ hai gồm có 304 điểm train tương ứng là 304 RPs, 40 điểm test tương ứng là 40 TPps và 4 điểm phát tín hiệu RSS tương ứng là 4 Aps.



Hình 3. 2 Đối tượng thực hiện thứ hai

Sai số các phương pháp:

Bảng 3. 3 Sai số tọa độ

Thuật toán	Sai số tọa độ (m)		
	Nhỏ nhất	Lớn nhất	Trung bình
KNN (304 điểm) k = 4	0.302	2.924	1.489
WKNN (304 điểm) k = 4	0.29	4.042	1.425
P&S-WKNN (304 điểm) k = 2	0	4.682	1.434

Trong quá trình thực hiện thuật toán P&S-WKNN, các tham số của mô hình suy hao có giá trị như sau:

$$n = 3.6; \text{ Pd0} = 36.6 \text{ dB}; \text{ d0} = 1 \text{ m}$$

Do tập dữ liệu train có 304 điểm, đã bao phủ không gian phòng. Vì vậy trong tập dữ liệu này không xét tới các phương pháp nội suy.

3.3 Nhận xét

Từ bảng 3.1 ta có thể thấy được sai số trung bình nhỏ nhất trong các phương pháp là 1.067m khi sử dụng thuật toán WKNN. Phần lớn sai số đều nằm dưới 1.2m, tuy nhiên có một vài điểm có sai số cao, một phần là do sai lệch trong việc lấy dữ liệu, phần nhiều là do sự thay đổi của môi trường giữa hai giai đoạn, cụ thể là trong giai đoạn trực tuyến, tại khu vực này có nhiều thiết bị phát sóng khác đang hoạt động khiến cho tín hiệu bị nhiễu khá lớn.

Bảng 3.2 ta có thể thấy được sai số RSSI của các thuật toán nội suy và ứng với thuật toán Gradient boosting là sai số nhỏ nhất với loss train là 7.947 (dB) và loss valid là 7.244 (dB). Việc nội suy này gây ra sai số tương đối là do việc nội suy ra giá trị RSS dựa theo các giá trị RSS trong tập train mà các giá trị này trong quá trình đo đạc đã có những sai số vì thế việc nội suy cũng gây ra sai số.

Bảng 3.3 cho thấy việc thiết lập một tập dữ liệu mới dựa trên đối tượng cũ có sai số cao hơn. Lý do đến từ việc đo đạc lại điểm dữ liệu không cùng điều kiện môi trường, dẫn tới sự chênh lệch này. Bên cạnh đó ta thấy sai số của phương pháp P&S-WKNN cải thiện hơn so với phương pháp KNN là 4%, và sai số của cả 3 phương pháp gần bằng nhau.

CHƯƠNG 4: KẾT LUẬN

4.1 Kết luận

Định vị sử dụng phương pháp Fingerprint với thuật toán Weigh K-nearest Neighbor đã có sai số nhỏ nhất trong các phương pháp đề ra. Tuy nhiên, việc giảm bớt ảnh hưởng từ nhiễu vẫn cần được điều chỉnh và lấy mẫu thêm.

Mô hình thực nghiệm đặt ở khu vực phòng Lab RF3i (C1 – 328) có nhiều điều kiện môi trường do nhiễu thiết bị có thể bật hoặc tắt ở các thời điểm nên việc hạn chế được nhiễu cần lấy mẫu tại nhiều thời điểm khác nhau trong ngày.

Ảnh hưởng bởi người chần cũng cần được lưu ý do trong quá trình đo có nhiều người di chuyển trong phòng.

Tuy có nhiễu lớn hơn so với đề ra, tuy nhiên, khi xét với các phương pháp sử dụng thuật toán khác thì phương pháp dấu vân tay vẫn đạt được sai số thấp hơn và ổn định hơn

4.2 Đề xuất các giải pháp khác

Bên cạnh các phương pháp đề ra, còn có nhiều yếu tố ảnh hưởng đến sai số mà có thể khắc phục được bằng cách kết hợp các phương pháp khác lại với nhau, hoặc sử dụng các bộ tiền xử lý trong pha thu thập dữ liệu để có thể lọc ra những giá trị lỗi hoặc giá trị bị lệch nhiều so với giá trị đúng của nó.

Đề xuất các phương pháp sau:

- Sử dụng bộ lọc Kalman Filter, Bayes Filter, Particle Filter
- Smallest Enclosing Circle-based Fingerprint
- C-means and Chi-square Distance
- Kết hợp với phương pháp Triangulation
- Nội suy sử dụng thuật toán Kriging

TÀI LIỆU THAM KHẢO

- [1] Z. Jianyong, L. Haiyong, C. Zili, and L. Zhaohui, “RSSI based Bluetooth low energy indoor positioning,” in 2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Busan, South Korea, Oct. 2014, pp. 526–533. doi: 10.1109/IPIN.2014.7275525.
- [2] Yapeng Wang, Xu Yang, Yutian Zhao, Yue Liu, and L. Cuthbert, “Bluetooth positioning using RSSI and triangulation methods,” in 2013 IEEE 10th Consumer Communications and Networking Conference (CCNC), Las Vegas, NV, Jan. 2013, pp. 837–842. doi: 10.1109/CCNC.2013.6488558.
- [3] S. Li and R. Rashidzadeh, “A Hybrid Indoor Location Positioning System,” in 2018 IEEE International Conference on Electro/Information Technology (EIT), Rochester, MI, May 2018, pp. 0187–0191. doi: 10.1109/EIT.2018.8500265.
- [4] S. G. Obreja and A. Vulpe, “Evaluation of an Indoor Localization Solution Based on Bluetooth Low Energy Beacons,” in 2020 13th International Conference on Communications (COMM), Bucharest, Romania, Jun. 2020, pp. 227–231. doi: 10.1109/COMM48946.2020.9141987.
- [5] R. Faragher and R. Harle, “Location Fingerprinting With Bluetooth Low Energy Beacons,” IEEE J. Select. Areas Commun., vol. 33, no. 11, pp. 2418–2428, Nov. 2015, doi: 10.1109/JSAC.2015.2430281.
- [6] Universitas Gadjah Mada Yogyakarta et al., “Integration of Indoor Localization System using Wi-Fi Fingerprint, Bluetooth Low Energy Beacon and Pedometer Based on Android Application Platform,” IJIES, vol. 13, no. 4, pp. 171–181, Aug. 2020, doi: 10.22266/ijies2020.0831.15.
- [7] Y. Zhuang, J. Yang, Y. Li, L. Qi, and N. El-Sheimy, “Smartphone-Based Indoor Localization with Bluetooth Low Energy Beacons,” Sensors, vol. 16, no. 5, p. 596, Apr. 2016, doi: 10.3390/s16050596.
- [8] L. Bai, F. Ciravegna, R. Bond, and M. Mulvenna, “A Low Cost Indoor Positioning System Using Bluetooth Low Energy,” IEEE Access, vol. 8, pp. 136858–136871, 2020, doi: 10.1109/ACCESS.2020.3012342.

- [9] P. Spachos, “Raw RSSI Dataset for Indoor Localization Fingerprinting.” IEEE DataPort, Apr. 17, 2020. doi: 10.21227/JRK5-QW26.
- [10] Bai, Y.B.; Wu, S.; Retscher, G.; Kealy, A.; Holden, L.; Tomko, M.; Borriak, A.; Hu, B.; Wu, H.R.; Zhang, K. A new method for improving Wi-Fi-based indoor positioning accuracy. *J. Locat. Based Serv.* 2014, 8, 135–147.
- [11] Shixiong Xia , Yi Liu , Guan Yuan ,*, Mingjun Zhu and Zhaohui Wang, “Indoor Fingerprint Positioning Based on Wi-Fi: An Overview,” *ISPRS Int. J. Geo-Inf.* 2017, 6, 135; doi:10.3390/ijgi6050135.
- [12] Xuesheng Peng , Ruizhi Chen , Kegen Yu , Feng Ye and Weixing Xue, “An Improved Weighted K-Nearest Neighbor Algorithm for Indoor Localization,” *Electronics* 2020, 9, 2117; doi:10.3390/electronics9122117.