

ỨNG DỤNG DEEP-LEARNING TRONG NHẬN DẠNG, XÁC ĐỊNH VỊ TRÍ CỦA ROBOT LÀM VIỆC TRONG NHÀ

Nguyễn Thế Anh, Hoàng Thị Nguyệt, Trần Anh Thảo

TÓM TẮT

Title: Application of Deep-Learning in recognition and localization of Indoor Robots

Từ khóa: nhận dạng robot, định vị robot, xử lý ảnh, học sâu, robot làm việc trong nhà, camera giám sát.

Keywords: robot recognition, robot localization, image processing, deep-learning, indoor robots, surveillance cameras.

Lịch sử bài báo

Ngày nhận bài: 21/03/2025

Ngày nhận kết quả bình duyệt: 24/04/2025

Ngày chấp nhận đăng bài: 14/05/2025

Tác giả: *Học viện Hải quân

Email liên hệ: nguyentheanh.uet@gmail.com

Việc nhận dạng và định vị robot là bài toán cốt lõi trong điều hướng robot trong nhà, đặt nền tảng cho các nhiệm vụ như lập kế hoạch quỹ đạo và tối ưu hóa di chuyển. Nhiều phương pháp đã được nghiên cứu, chẳng hạn như so sánh dữ liệu quét laser với bản đồ, sử dụng thẻ RFID, mã QR, hoặc hệ thống iGPS. Trong bài báo này, chúng tôi đề xuất một giải pháp chi phí thấp sử dụng camera giám sát để nhận dạng và xác định vị trí robot trong không gian rộng như sảnh lớn tại sân bay, bến tàu, v.v. Phương pháp kết hợp học sâu dựa trên mô hình SSD_MobileNet_v2 để nhận dạng robot và xử lý ảnh nhằm tính toán khoảng cách từ camera đến robot thông qua đối tượng nhận dạng gắn trên robot. Kết quả thực nghiệm cho thấy tỷ lệ nhận dạng chính xác đạt 96% và sai số vị trí tối đa là 150 mm, khẳng định tính khả thi của giải pháp trong môi trường thực tế.

ABSTRACT

The recognition and localization of robots represent a core challenge in indoor robot navigation, laying the foundation for tasks such as trajectory planning and movement optimization. Numerous methods have been explored, including laser scan data comparison with maps, the use of RFID tags, QR codes, or iGPS systems. In this paper, we propose a low-cost solution utilizing surveillance cameras to recognize and determine the position of robots in expansive indoor spaces, such as large halls in airports, piers, etc. The approach integrates deep learning based on the SSD_MobileNet_v2 model for robot recognition and image processing to calculate the distance from the camera to the robot via an identification object mounted on it. Experimental results demonstrate a recognition accuracy of 96% and a maximum localization error of 150 mm, confirming the feasibility of the solution in real-world environments.

1. Đặt vấn đề/Giới thiệu

Trong những năm gần đây, các loại robot làm việc trong nhà đã trở thành một lĩnh vực nghiên cứu và phát triển quan trọng, với mục tiêu hỗ trợ con người trong việc giảm bớt công việc thủ công, nâng cao

hiệu quả lao động và tiết kiệm thời gian. Các ứng dụng thực tiễn của robot làm việc trong nhà bao gồm vận chuyển hàng hóa trong kho, cung cấp dịch vụ trong khách sạn, bệnh viện, hoặc thậm chí hỗ trợ phục vụ tại các khu vực công cộng như sân bay, bến tàu. Một

trong những hướng nghiên cứu được các nhà khoa học đặc biệt quan tâm là phát triển các dòng robot có khả năng tự động hóa cao, chẳng hạn như robot vận chuyển hàng hóa tự hành hoặc robot dịch vụ thông minh. Tuy nhiên, để những robot này có thể hoạt động hiệu quả trong môi trường trong nhà, vấn đề đầu tiên và quan trọng nhất cần giải quyết là khả năng nhận dạng và xác định vị trí chính xác của robot trong không gian làm việc. Đây là bước cơ bản, đặt nền tảng cho các nhiệm vụ tiếp theo như lập kế hoạch quỹ đạo di chuyển, tối ưu hóa đường đi, hay tương tác với môi trường xung quanh.

Để giải quyết bài toán nhận dạng và định vị robot, nhiều phương pháp đã được các nhà nghiên cứu đề xuất và triển khai trong thời gian qua. Chẳng hạn, phương pháp so sánh dữ liệu quét laser với bản đồ (Chou & Liu, 2013) tận dụng khả năng quét không gian của công nghệ laser để xây dựng bản đồ môi trường, từ đó xác định vị trí robot dựa trên dữ liệu thu thập được. Một cách tiếp cận khác là sử dụng thẻ RFID hoặc mã QR (Xiao-Long et al., 2017), trong đó các thẻ hoặc mã được bố trí tại các điểm cố định để robot đọc và xác định vị trí của mình. Hệ thống iGPS (Norman et al., 2010) cung cấp một giải pháp định vị dựa trên tín hiệu hồng ngoại, trong khi hệ thống định vị quán tính (INS) kết hợp với cảm biến IMU (Shamseldin et al., 2018) sử dụng gia tốc kế và con quay hồi chuyển để theo dõi chuyển động của robot từ điểm xuất phát. Ngoài ra, công nghệ LIDAR (Gallant & Marshall, 2016) với khả năng quét 3D độ chính xác cao cũng được ứng dụng rộng rãi trong việc định vị robot. Tuy nhiên, các phương pháp này thường gặp phải một số hạn chế khi triển khai thực tế. Chẳng hạn, công nghệ LIDAR và cảm biến IMU chất lượng cao đòi hỏi chi phí đầu tư lớn cho thiết bị, trong khi phương pháp sử dụng thẻ RFID hoặc mã QR yêu cầu lắp đặt thêm

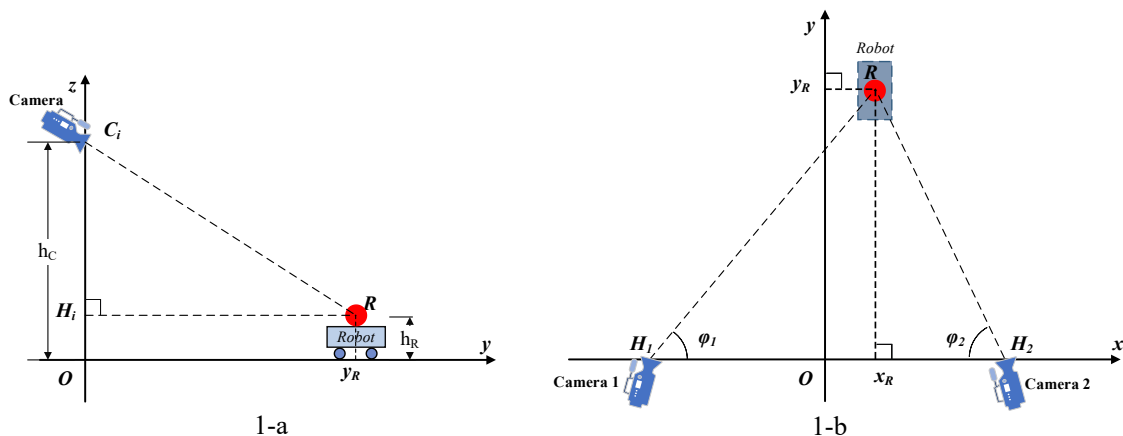
cơ sở hạ tầng như các thẻ định vị tại nhiều vị trí trong không gian làm việc. Phương pháp quét laser lại phụ thuộc vào việc xây dựng bản đồ chi tiết trước đó, đòi hỏi thời gian và công sức chuẩn bị, chưa kể hệ thống iGPS cần thiết lập mạng lưới tín hiệu phức tạp, gây khó khăn khi áp dụng trong các khu vực rộng lớn như sân chờ sân bay hay bến tàu.

Trước những thách thức đó, trong bài báo này, chúng tôi đề xuất một giải pháp chi phí thấp nhằm nhận dạng và xác định vị trí robot làm việc trong nhà, tận dụng các camera giám sát sẵn có tại những địa điểm như sảnh lớn của sân bay hoặc bến tàu. Phương pháp này không chỉ giảm thiểu sự phụ thuộc vào thiết bị chuyên dụng đắt tiền mà còn đơn giản hóa quá trình triển khai trong thực tế. Cụ thể, chúng tôi kết hợp học sâu dựa trên mô hình SSD_MobileNet_v2 (TensorFlowGardener, 2021) để nhận dạng robot một cách nhanh chóng và chính xác, đồng thời sử dụng xử lý ảnh để tính toán khoảng cách từ camera đến robot thông qua một đối tượng nhận dạng gắn trên robot. Giải pháp này hướng tới việc cung cấp một phương thức hiệu quả, dễ tiếp cận và linh hoạt cho việc điều hướng robot làm việc trong nhà, đặc biệt trong các không gian rộng lớn và phức tạp, mở ra tiềm năng ứng dụng thực tiễn trong nhiều lĩnh vực.

2. Giải pháp nhận dạng và định vị robot trong không gian làm việc

2.1. Phương pháp xác định tọa độ robot trong không gian

Để xác định vị trí robot trong không gian, ta sử dụng 2 camera như Hình 1. Trong đó Oxy là mặt phẳng sàn còn Oyz là mặt phẳng thẳng đứng vuông góc với Oxy. Vị trí của 2 camera, chiều cao camera so với sàn và khoảng cách giữa 2 camera được đảm bảo sao cho robot luôn nằm trong trường quan sát của 2 camera.



Hình 1. Hình chiếu của robot trong không gian làm việc

Trên robot, đặt một đối tượng được nhận dạng (một quả bóng màu xanh trong Hình 1). Đối tượng được nhận dạng được đặt trên robot, vì vậy nó di chuyển cùng với chuyển động của robot. Chiều cao h_R thể hiện chiều cao của đối tượng được nhận dạng so với mặt phẳng chuyển động của robot. Nếu xác định được khoảng cách giữa camera và đối tượng được nhận dạng (RC_1 và RC_2), ta hoàn toàn có thể xác định độ dài hình chiếu của RC_1 và RC_2 trên sàn là RH_1 và RH_2 tương ứng. Từ đó xác định được tọa độ x_R và y_R của robot.

Phương pháp nhận dạng robot và xác định khoảng cách RC_1 và RC_2 được trình bày trong Mục 2.2, sử dụng phương pháp học sâu để nhận dạng robot và phương pháp xấp xỉ hàm đa thức để xác định khoảng cách của camera và đối tượng được nhận dạng.

Theo Hình 1-b, tọa độ của robot (x_R, y_R) được xác định như sau:

$$x_R = RH_1 \cos \varphi_1 - OH_1, \quad y_R = RH_1 \sin \varphi_1 \quad (1)$$

Mặt khác theo Hình 1-a (với $i=1$ tương ứng Camera 1 và $i=2$ tương ứng Camera 2):

$$RH_1 = \sqrt{RC_1^2 - (h_C - h_R)^2}, \quad RH_2 = \sqrt{RC_2^2 - (h_C - h_R)^2} \quad (2)$$

Với 2 camera được bố trí quan sát robot như Hình 1-b nên tọa độ y_R của robot luôn dương nên ta có:

$$\cos \varphi_1 = \frac{(RH_1)^2 + (H_1 H_2)^2 - (RH_2)^2}{2(RH_1)(H_1 H_2)}$$

$$\sin \varphi_1 = \sqrt{1 - \left[\frac{(RH_1)^2 + (H_1 H_2)^2 - (RH_2)^2}{2(RH_1)(H_1 H_2)} \right]^2} \quad (3)$$

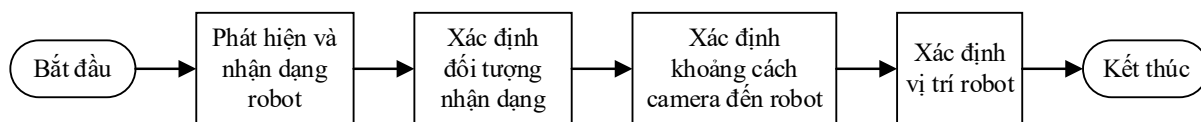
Thay (2), (3) vào (1) ta tính được tọa độ của robot:

$$x_R = \frac{(RC_1)^2 + (H_1 H_2)^2 - (RC_2)^2}{2(H_1 H_2)} - OH_1$$

$$y_R = \frac{\sqrt{4(H_1 H_2)^2 [(RC_1)^2 - (h_C - h_R)^2] - [(RC_1)^2 + (H_1 H_2)^2 - (RC_2)^2]^2}}{2(H_1 H_2)} \quad (4)$$

2.2. Xây dựng các bước nhận dạng và xác định vị trí của robot

Từ phương pháp đã trình bày ở Mục 2.1, có thể đưa ra các bước tổng quát để xác định vị trí của robot trong không gian làm việc của nó như mô tả trong Hình 2.



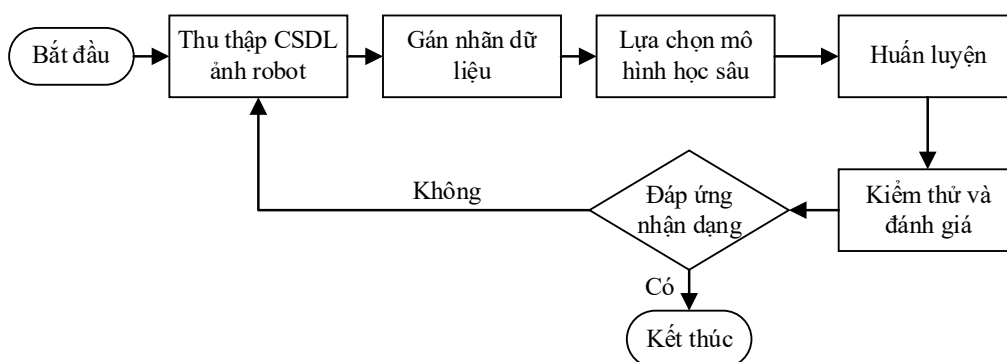
Hình 2. Các bước xác định vị trí của robot

Bước 1: Phát hiện và nhận dạng robot.

Hệ thống 2 camera (camera 1 và camera 2) đóng vai trò thu nhận hình ảnh không gian hoạt động của robot theo thời gian thực, từ đó chương trình sẽ phát hiện và nhận dạng robot trong không gian làm việc.

Hiện nay nhiều phương pháp được sử dụng để phát hiện và nhận dạng robot trong đó phương pháp học sâu (deep-learning) nổi bật với khả năng nhận dạng nhanh và độ

chính xác cao. Nhóm tác giả lựa chọn phương án sử dụng bộ thư viện Tensorflow và mô hình SSD_mobilenet_v2 (Single Shot MultiBox Detector) có tốc độ xử lý nhanh nhưng vẫn đảm bảo độ chính xác để huấn luyện mạng (tốc độ xử lý 27ms với ảnh độ phân giải 600×600 pixel (TensorFlowGardener, 2021), phù hợp với ứng dụng theo dõi đối tượng theo thời gian thực. Các bước tổng quát để thực hiện huấn luyện được mô tả trong Hình 3.



Hình 3. Các bước tổng quát phát hiện và nhận dạng robot

Bước 2: Xác định đối tượng cần nhận dạng trên robot

Chương trình sẽ thực hiện xác định đối tượng nhận dạng là quả bóng gắn trên robot. Nếu đúng là quả bóng có màu xanh (đúng về hình dáng, màu sắc, kích thước) thì chương trình sẽ cung cấp các thông số về kích thước đường kính của quả bóng (tính bằng pixel).

Bước 3: Xác định khoảng cách từ camera đến robot

Sau khi xác định được kích thước đường kính của bóng, chương trình sẽ chuyển sang tính toán khoảng cách từ các

camera đến đối tượng nhận dạng (D_1 và D_2). Khoảng cách này cũng chính là khoảng cách từ camera đến robot. Khoảng cách từ camera đến vật thể nhận dạng (đặt trên robot) được tính theo công thức (5).

Hiện nay, có nhiều phương pháp khác nhau để xác định khoảng cách từ camera đến đối tượng được nhận diện. Mỗi phương pháp có ưu điểm và nhược điểm của nó. Trong phương pháp sử dụng camera stereo (Krishnan & Kollipara, 2014), ưu điểm là độ chính xác cao, nhưng hệ thống camera này có giá thành rất cao. Phương pháp sử dụng camera kết hợp với cảm biến con quay hồi

chuyển cũng cho kết quả tương đối chính xác, tuy nhiên nhược điểm của nó là giá thành hệ thống còn cao do sử dụng cảm biến con quay hồi chuyển và quá trình thiết lập tính toán phức tạp (Satoh et al., 2004).

Để khắc phục nhược điểm của các phương pháp trên, tác giả đề xuất sử dụng phương pháp thực nghiệm. Bằng cách đặt quả bóng ở các vị trí và khoảng cách cố định, chúng ta có thể xác định được kích thước đường kính quả bóng bằng pixel trong hình ảnh thu được. Lặp

lại nhiều lần đo ở các khoảng cách khác nhau, ta nhận được một tập dữ liệu đầu vào. Từ đó, xây dựng hàm tương quan giữa kích thước của quả bóng trên ảnh với khoảng cách của quả bóng tới camera.

Với các thông số thực nghiệm như sau: đường kính bóng là 110 mm, độ phân giải của camera là 1280×720 pixel. Ta xây dựng được bảng tương quan giữa kích thước của bóng trên ảnh thu được và khoảng cách thực của bóng tới camera (Bảng 1).

Bảng 1. Bảng tương quan giữa kích thước quả bóng và khoảng cách của nó đến camera

Thứ tự thực nghiệm	1	2	3	4	5	6	7	8	9	10	11	12
Khoảng cách thực từ camera đến quả bóng (cm)	50	80	110	140	170	200	230	260	290	320	350	380
Đường kính bóng trên ảnh (pixel)	344	296	210	130	100	80	70	58	52	48	44	42

Áp dụng phương pháp nội suy và xấp xỉ hàm đa thức, chúng ta tìm được hàm tương quan bậc 4 giữa kích thước của quả bóng trên ảnh và khoảng cách giữa quả bóng và camera.

$$RC_i = 2.7288 \cdot 10^{-7} p^4 - 2.5214 \cdot 10^{-4} p^3 + 0.0824 p^2 - \quad (5)$$

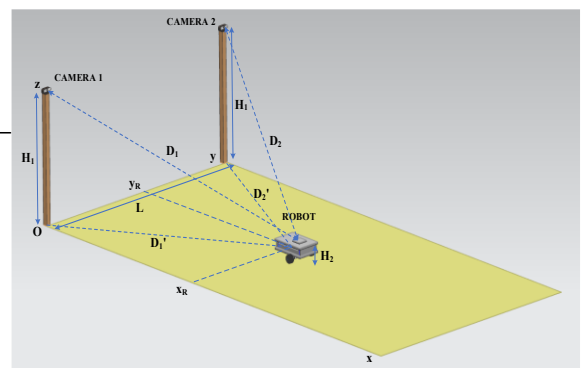
Trong đó: p là đường kính của bóng đo được trên ảnh (pixel); RC_i là khoảng cách thực tế từ quả bóng đến camera.

Bước 4: Xác định vị trí của robot

Sau khi tính được khoảng cách của robot đến camera theo công thức (5), áp dụng công thức (4) ta tính được tọa độ x_R và y_R của robot trong mặt phẳng Oxy gắn với không gian làm việc của nó.

3. Thực nghiệm

3.1. Xây dựng môi trường thực nghiệm



Hình 4. Môi trường thực nghiệm hoạt động của robot

Để tiến hành thí nghiệm, nhóm tác giả đã thiết lập môi trường hoạt động theo Hình 4 với các thông số cụ thể như sau:

- Kích thước vùng di chuyển của robot: 200 x 500 (cm);

- Chiều cao của camera so với mặt sàn:

$$H_1 = H_2 = 100\text{cm}$$

- Khoảng cách giữa 2 camera: $L = 200\text{cm}$;

- Xác định gốc tọa độ Oxy, sao cho:

+ Vị trí đặt camera 1 ứng với tọa độ:

$$x_{cam1} = 0\text{ cm}; y_{cam1} = 0\text{ cm};$$

+ Vị trí đặt camera 2 ứng với tọa độ:

$$y_{cam2} = 200\text{ cm}; x_{cam2} = 0\text{ cm}.$$

Ngoài ra, trong không gian hoạt động của robot, ta đánh dấu các vị trí với tọa độ được xác định trước. Các vị trí này sẽ dùng để kiểm tra độ chính xác của chương trình tính vị trí của robot so với vị trí thực tế.

3.2. Thực nghiệm và đánh giá sai số

Để thực nghiệm, đánh giá sai số nhận dạng của robot và thuật toán xác định vị trí của robot, nhóm tác giả sử dụng Google Colab để huấn luyện nhận dạng và xây dựng chương trình tính toán vị trí của robot trên phần mềm LabVIEW.

3.2.1. Thực nghiệm và đánh giá sai số nhận dạng robot

Trong thí nghiệm này, đầu tiên nhóm tác giả xây dựng cơ sở dữ liệu huấn luyện gồm 275 ảnh chụp khác nhau của robot. Sau khi huấn luyện mô hình nhận dạng từ cơ sở dữ liệu ảnh trên với 4033 bước huấn luyện, tác giả sử dụng 25 ảnh robot khác được chụp từ camera với các vị trí và các góc khác nhau để kiểm tra. Kết quả thu được biểu diễn trên Bảng 2 với 24/25 ảnh (tỷ lệ 96%) được nhận dạng chính xác với độ tin cậy dự đoán đạt trên 0.96 (độ tin cậy dự đoán có giá trị từ 0 đến 1, càng gần 1 thì độ khẳng định chính xác đối tượng cần nhận dạng càng cao). Ảnh còn lại (ảnh số 18 - hình 5) tuy robot vẫn nằm trong khung hình nhận dạng nhưng do khung bao quá lớn (bao gồm cả những vùng không phải là ảnh robot) nên vẫn đánh giá là nhận dạng không chính xác.

Bảng 2. Bảng kết quả nhận dạng robot

TT Ảnh	Nhận dạng đúng	Độ tin cậy dự đoán	TT Ảnh	Nhận dạng đúng	Độ tin cậy dự đoán	TT Ảnh	Nhận dạng đúng	Độ tin cậy dự đoán
1	Có	1	10	Có	1	18	Không	1
2	Có	1	11	Có	1	19	Có	1
3	Có	1	12	Có	1	20	Có	0.98
4	Có	1	13	Có	1	21	Có	1
5	Có	1	14	Có	1	22	Có	1
6	Có	1	15	Có	0.97	23	Có	0.96
7	Có	1	16	Có	1	24	Có	0.98
8	Có	1	17	Có	1	25	Có	1
9	Có	1						

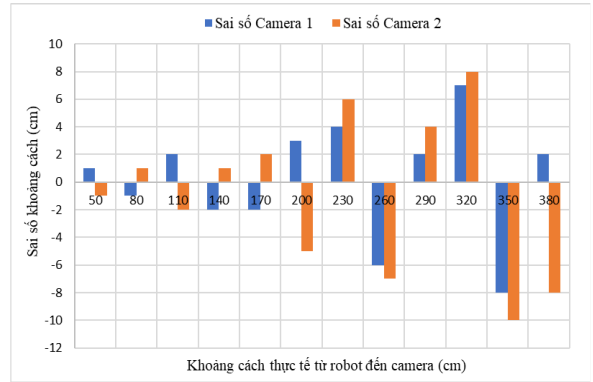


Hình 5. Kết quả nhận dạng robot trong ảnh số 18

3.2.2. Thực nghiệm và đánh giá sai số khoảng cách từ camera đến robot

Trong thí nghiệm này, nhóm tác giả tiến hành kiểm tra độ chính xác của hàm xấp xỉ (5) để xác định khoảng cách từ camera tới robot. Các bước thực nghiệm như sau: di chuyển robot đến các vị trí cho trước trên sàn (khoảng cách các vị trí này tới camera đã biết). Sau đó, so sánh kết quả tính toán từ chương trình và dữ liệu thực tế.

Sai số khoảng cách của camera 1 và camera 2 tới robot được biểu diễn trên Hình 5. Ta quan sát thấy sai số lớn nhất của ảnh của camera 1 là 80mm. Sai số lớn nhất của camera 2 là 100mm. Robot di chuyển khỏi camera càng xa thì sai số tính toán khoảng cách càng lớn.



Hình 5. Kết quả thí nghiệm đo khoảng cách từ robot đến camera

3.2.3. Thực nghiệm và đánh giá sai số vị trí của robot

Trong thí nghiệm này, nhóm tác giả sẽ kiểm tra sai số vị trí (sai số tổng thể σ_R) của robot so với vị trí thực được tính toán dựa trên hai trường hợp: (i) sai số theo tọa độ x_R (sai số σ_x), và (ii) sai số theo tọa độ y_R (sai số σ_y).

$$\sigma_R = \sqrt{\sigma_x^2 + \sigma_y^2} \quad (6)$$

Kết quả thực nghiệm được biểu diễn trong Bảng 3. Ta quan sát thấy sai số tối đa là 150mm. Sai số có xu hướng tăng lên khi robot di chuyển xa hơn khỏi trường giám sát của hai camera.

Bảng 3. Bảng sai số vị trí của robot

$Y_R(\text{cm}) \backslash X_R(\text{cm})$	50	80	110	140	170	200	230	260	290	320	350	380
0	10	8	7	7	7	6	7	11	12	13	14	14
20	8	7	6	6	5	6	7	9	11	12	14	14
40	5	6	4	4	3	4	5	5	7	8	13	13
60	3	4	5	2	2	3	4	4	5	7	11	12
80	2	4	4	1	2	1	2	3	5	6	9	10
100	2	2	1	0	0	0	2	3	5	6	9	10
120	3	2	1	0	1	2	2	2	6	6	10	11
140	4	3	3	1	1	3	3	3	6	7	11	12
160	5	4	3	3	2	4	5	7	8	10	13	14
180	7	7	6	5	4	6	7	8	10	12	14	15
200	11	10	9	8	7	8	9	9	11	13	14	15

4. Kết luận

Bài báo trình bày ứng dụng phương pháp học sâu trên bộ thư viện Tensorflow để nhận dạng robot và giải pháp ứng dụng xử lý ảnh để xây dựng hàm tương quan khoảng cách giữa camera quan sát và robot thông qua đối tượng nhận dạng (đặt trên robot). Qua đó, xây dựng chương trình tính toán vị trí (X_R , Y_R) của robot trong không gian làm việc. Kết quả thực nghiệm thể hiện kết quả tốt của chương trình nhận dạng và chương trình tính toán vị trí (khả năng nhận dạng chính xác 96%, sai số vị trí lớn nhất là 150mm). Nguyên nhân nhận dạng nhầm là cơ sở dữ liệu ít, số bước huấn luyện và

đặc trưng hình ảnh robot thấp. Trường hợp sai số vị trí, nguyên nhân là chất lượng thu nhận hình ảnh của camera bị thay đổi khi robot di chuyển ra xa camera. Số lượng pixel của đối tượng được nhận dạng càng ít thì sai số tính toán khoảng cách từ camera đến robot càng lớn. Để nâng cao độ chính xác vị trí, cần tăng độ phân giải của camera quan sát.

Trong các nghiên cứu tiếp theo, nhóm tác giả sẽ tập trung vào cải thiện độ chính xác của việc nhận dạng và thuật toán xác định vị trí của robot sau đó tiến hành nghiên cứu sử dụng camera giám sát để lập quỹ đạo cho robot di chuyển trong không gian làm việc.

TÀI LIỆU THAM KHẢO

- Chou, Y.-S., & Liu, J.-S. (2013). A robotic indoor 3D mapping system using a 2D laser range finder mounted on a rotating four-bar linkage of a mobile platform. *International Journal of Advanced Robotic Systems*, 10(1), 45.
- Gallant, M. J., & Marshall, J. A. (2016). Two-dimensional axis mapping using LiDAR. *IEEE Transactions on Robotics*, 32(1), 150-160.
- Krishnan, A. B., & Kollipara, J. (2014). Intelligent indoor mobile robot navigation using stereo vision. *arXiv preprint arXiv:1412.6153*.
- Norman, A., Schönberg, A., Gorlach, I., & Schmitt, R. (2010). Cooperation of industrial robots with indoor-GPS. *Proceedings of the international conference on competitive manufacturing*, 215-224.
- Satoh, K., Uchiyama, S., & Yamamoto, H. (2004). A head tracking method using bird's-eye view camera and gyroscope. *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*, 202-211.
- Shamseldin, T., Manerikar, A., Elbahnasawy, M., & Habib, A. (2018). SLAM-based Pseudo-GNSS/INS localization system for indoor LiDAR mobile mapping systems. *2018 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, 197-208.
- TensorFlowGardener. (2021). *TensorFlow 1 Detection Model Zoo*. tensorflow. https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/tf1_detection_zoo.md
- Xiao-Long, W., Chun-Fu, W., Guo-Dong, L., & Qing-Xie, C. (2017). A robot navigation method based on RFID and QR code in the warehouse. *2017 Chinese automation congress (CAC)*, 7837-7840.