

Điều khiển bám người đồng thời tránh vật cản cho robot di động bằng công nghệ xử lý ảnh dựa trên kỹ thuật học sâu

Human following and collision avoidance control of mobile robots by vision-based deep neural network

Bùi Trung Nghĩa, Nguyễn Văn Nam, Nguyễn Duy Phương, Nguyễn Công Minh, Dương Văn Đạt, Vũ Nhật Cường¹ và Nguyễn Mạnh Linh^{1*}

¹Đại học Bách khoa Hà Nội

*Corresponding author email: linh.nguyenmanh@hust.edu.vn

Abstract

Nowadays, mobile robots have been popular not only in industrial applications such as materials transportation but also in non-industrial applications, e.g., human assistance. Among developed configurations, omnidirectional mobile robots have attracted great attention recently due to their superior maneuverability over their conventional counterparts. In this research, an application of a four mecanum-wheeled omnidirectional mobile robot (4-MWMMR) in human assistance has been developed. By using a vision-based deep neural network in real-time, the 4-MWMMR is capable of following an authorized person, collision avoidance and obeying the hand pose command, thereby assisting users in transporting materials in unknown environment. Good experimental results show the ability of the developed system to be used in practice.

Keywords: Omnidirectional mobile robot, Vision-based deep neural network, Convolution neural network.

Các từ viết tắt

OMR	Omnidirectional mobile robot
4-MWMMR	4-Mecanum wheeled mobile robot
SSD	Single shot detector
VGG	Visual graphic group
CNN	Convolution neural network
PCA	Principle component analysis

Tóm tắt

Ngày nay, robot di động đã trở nên phổ biến không những trong công nghiệp, chẳng hạn như vận chuyển vật tư giữa các công đoạn sản xuất, mà còn trong dân dụng như hỗ trợ con người trong cơ sở y tế, siêu thị, kho hàng. Trong số các cấu hình đã được phát triển, robot di động sử dụng bánh xe đa hướng đặc biệt được chú ý trong thời gian gần đây, do khả năng cơ động vượt trội so với các đối thủ sử dụng bánh xe truyền thống. Trong nghiên cứu này, một robot di động sử dụng cấu hình bốn bánh đa hướng kiểu Mecanum (4-MWMMR) với khả năng tương tác và hỗ trợ con người được phát triển. Bằng cách xử lý hình ảnh trong thời gian thực dựa trên kỹ thuật học sâu, 4-MWMMR có khả năng di chuyển bám theo người được cấp quyền đồng thời tránh chướng ngại vật, cũng như có thể được điều khiển từ xa qua cử chỉ tay, từ đó hỗ

trợ người dùng trong ứng dụng vận chuyển vật tư hàng hóa trong môi trường không có bản đồ lập sẵn. Các kết quả thử nghiệm bước đầu cho thấy tiềm năng của hệ thống có thể được ứng dụng trong thực tế.

1. Giới thiệu chung

Gần đây, ứng dụng robot di động bùng nổ vì tính cơ động và hiệu quả cao. Nhiều nguyên mẫu và sản phẩm khác nhau đã được phát triển để đáp ứng các yêu cầu khắt khe của người dùng. Với sự phát triển mạnh của các kỹ thuật học sâu, khả năng tích hợp cao về sự nhận biết, điều hướng và tương tác với người dùng được đặc biệt quan tâm trong thời gian gần đây [1]. Thông thường, các robot di động sử dụng bánh xe tròn truyền thống với các cấu hình phổ biến như hai bánh vi sai hoặc bốn bánh giống ô tô. Một nhược điểm rõ ràng của các cấu hình nêu trên là chúng không có khả năng chuyển hướng tức thời mà cần thực hiện từ từ thông qua bánh lái, do đó hạn chế khả năng di chuyển trong không gian hẹp. Để khắc phục nhược điểm này, robot di động đa hướng (OMR) với bánh chuyên dụng đã được phát triển, mang lại khả năng cơ động cao hơn cũng như thích ứng tốt trong không gian hẹp. Hai cấu hình OMR được sử dụng rộng rãi trong nghiên cứu cũng như ứng dụng thực tế là 3 bánh và 4 bánh. Cấu hình 3 bánh thường được sử dụng để thiết kế robot di động nhỏ với trọng tải nhẹ. Trong trường hợp tải trọng

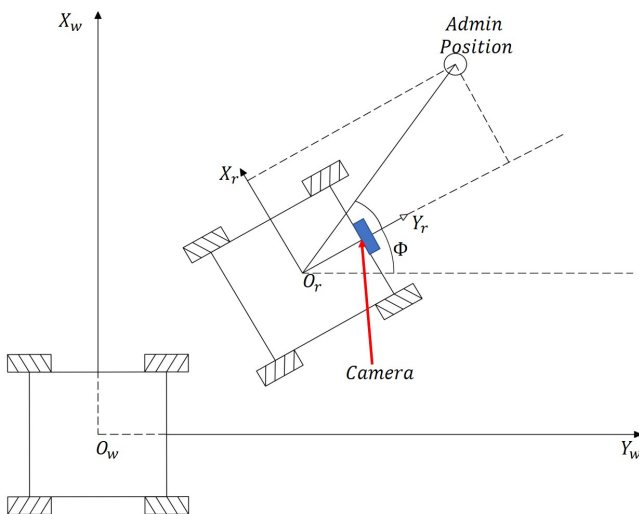
nặng, cấu hình 4 bánh thường được sử dụng [2].

Nhìn chung, hệ thống điều khiển của OMR có thể được phân loại thành điều khiển cấp chấp hành, tập trung vào bài toán động học và động lực học của robot, cũng như xử lý các vấn đề liên quan đến an toàn vận hành như tránh va chạm. Và hệ điều khiển cấp cao, chủ yếu xử lý các nhiệm vụ điều hướng dựa trên dữ liệu camera và cảm biến laser. Với sự phát triển bùng nổ của trí tuệ nhân tạo (AI) trong thập kỷ này, kết hợp với cảm biến hình ảnh, lượng thông tin thu thập đủ dùng cho hầu hết các vấn đề liên quan đến điều khiển robot di động, nhờ đó không chỉ nâng cao khả năng tương tác với người dùng mà còn làm cho robot ngày càng thông minh hơn. Do đó, các hệ thống điều khiển cấp cao dựa trên AI trở thành đề tài hấp dẫn trong thời gian gần đây [3].

Để đảm bảo robot di động có thể di chuyển và thực hiện các nhiệm vụ của chúng trong môi trường phức tạp, hầu hết các nhà nghiên cứu tập trung vào giải quyết vấn đề lập bản đồ và định vị [4, 5] mà ít tập trung vào vấn đề tương tác giữa robot di động và người sử dụng. Trong nghiên cứu này, một robot di động với cấu hình bốn bánh đa hướng mecanum được phát triển với mục tiêu chính là tăng khả năng tương tác với người sử dụng. Các tính năng mới được phát triển bao gồm: nhận dạng khuôn mặt của người được cấp quyền (admin), bấm theo admin hoặc nhận lệnh bằng cử chỉ tay của admin để thực hiện các thao tác di chuyển, đồng thời có khả năng tránh các chướng ngại vật gặp phải trên đường đi. Các tính năng rất hữu ích trong trường hợp cần robot di chuyển ngoài quỹ đạo đã thiết lập từ trước trên mặt bằng sản xuất (line từ, vạch màu), hoặc trong môi trường không biết trước bản đồ như trong siêu thị, kho hàng, cơ sở y tế ... Đã có một số nghiên cứu về robot di động tương tác với con người dựa trên xử lý ảnh [6, 7]. Tuy nhiên trong nghiên cứu này, nhóm tác giả tập trung vào việc sử dụng những thành tựu mới của mạng neural học sâu (deep neural network) trong xử lý ảnh để giải quyết bài toán trên.

2. Mô tả hệ thống

2.1. Mô hình động học của 4-MWMR



Hình 1. Hệ tọa độ của mobile robot

Cấu hình của robot di động và các khung tọa độ tương ứng được minh họa ở hình 1, trong đó X_w, O_w, Y_w và X_r, O_r, Y_r lần lượt là hệ tọa độ gốc và hệ tọa độ gắn với thân xe. Đặt $[\theta_1, \theta_2, \theta_3, \theta_4]^T$

Bảng 1. Tham số của mô hình mobile robot

Ký hiệu	Mô tả	Giá trị	Đơn vị
W	Một nửa chiều rộng xe	0.3	m
L	Một nửa chiều dài xe	0.3	m
m	Tổng khối lượng xe	40	kg
R	Bán kính bánh xe	0.076	m
$\dot{\theta}_i$	Vận tốc góc bánh xe thứ i	-	rad/s
Φ	Góc lệch giữa 2 hệ tọa độ	-	rad

và $[\dot{x}_r, \dot{y}_r, \dot{\Phi}]^T$ lần lượt là vector vận tốc góc của bánh xe và vận tốc xe trên hệ tọa độ thân xe, khi đó ta có mối quan hệ động học thuận và ngược lần lượt được biểu diễn bởi (1) và (2) như sau:

$$[\dot{\theta}_1 \quad \dot{\theta}_2 \quad \dot{\theta}_3 \quad \dot{\theta}_4]^T = \frac{1}{R} J [\dot{x}_r \quad \dot{y}_r \quad \dot{\Phi}]^T \quad (1)$$

$$[\dot{x}_r \quad \dot{y}_r \quad \dot{\Phi}]^T = R J^+ [\dot{\theta}_1 \quad \dot{\theta}_2 \quad \dot{\theta}_3 \quad \dot{\theta}_4]^T \quad (2)$$

Trong đó,

$$J = \begin{bmatrix} 1 & 1 & L+W \\ -1 & 1 & -L+W \\ 1 & 1 & -L+W \\ -1 & 1 & L+W \end{bmatrix} \quad (3)$$

và $J^+ = (J^T J)^{-1} J^T$ là ma trận giả nghịch đảo của J :

$$J^+ = \frac{1}{4} \begin{bmatrix} 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \\ \frac{1}{L+W} & -\frac{1}{L+W} & -\frac{1}{L+W} & \frac{1}{L+W} \end{bmatrix} \quad (4)$$

Các thông số của mô hình được cung cấp chi tiết ở bảng 1. Còn tốc độ dịch chuyển của thân xe quy đổi về hệ tọa độ gốc được tính bởi phép chuyển đổi (5):

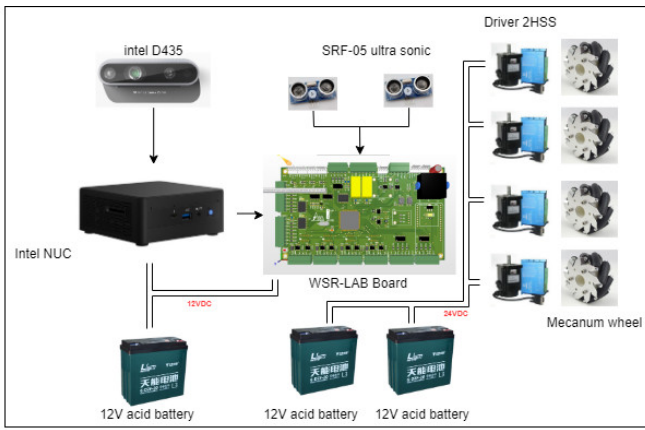
$$[\dot{x}_w \quad \dot{y}_w \quad \dot{\Phi}_w]^T = \mathfrak{R}(\Phi) [\dot{x}_r \quad \dot{y}_r \quad \dot{\Phi}_r]^T \quad (5)$$

trong đó $\mathfrak{R}(\Phi)$ là ma trận chuyển tọa độ biểu diễn bởi (6):

$$\mathfrak{R}(\Phi) = \begin{bmatrix} \cos(\Phi) & -\sin(\Phi) & 0 \\ \sin(\Phi) & \cos(\Phi) & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (6)$$

2.2. Trang bị điện cho 4-MWMR

Để thực hiện các chức năng điều khiển như đã đề cập ở trên, hệ thống điều khiển xe được thiết kế với kiến trúc như mô tả ở hình 2. Có thể chia hệ thống điều khiển ra hai tầng. Hệ điều khiển trung tâm được trang bị máy tính hiệu suất cao Intel NUC-11 NUC11TNH170Z với bộ xử lý Intel core i7 thế hệ thứ 11 để có thể thực hiện các thuật toán xử lý ảnh trong thời gian thực. Các thông tin về đối tượng và môi trường xung quanh được thu thập và gửi đến hệ thống điều khiển bởi máy ảnh cảm biến độ sâu intel realsense D435i. Ngoài các thông số cơ bản như độ phân giải 1280x720, tốc độ xử lý tới 90 khung hình/giây, D435i còn cung cấp các thông tin quan trọng khác cho điều hướng như khoảng cách đến các đối tượng trong dải từ 0.1m đến 10m, gia tốc và vận tốc góc của phần khung robot di động thông qua cảm biến đo lường quán tính (IMU) tích hợp trên thiết bị.

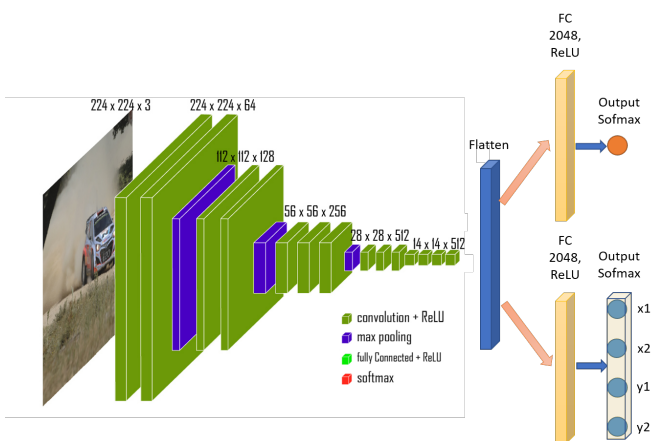


Hình 2. Trang bị điện của robot di động đa hướng

Ở tầng điều khiển cấp thấp hơn, mỗi bánh xe được truyền động bởi một bộ servo sử dụng động cơ bước lai. Đây là loại động cơ đặc biệt phù hợp cho các ứng dụng có tốc độ quay thấp nhưng đòi hỏi mô men lớn. Việc đảm bảo an toàn và chống va chạm cho xe trong quá trình vận hành được thực hiện bởi các cảm biến siêu âm lắp ở 4 phía của xe. Trung tâm của hệ điều khiển cấp này là một bảng mạch do nhóm nghiên cứu tự phát triển trên nền tảng vi điều khiển 32 bit STM32F103ZCT6. Bảng mạch điều khiển nhúng này đóng vai trò cầu nối giữa máy tính điều khiển trung tâm và các thiết bị truyền động, đồng thời xử lý trực tiếp các tín hiệu liên quan đến an toàn trong quá trình vận hành như tín hiệu từ cảm biến va chạm kiểu tiếp xúc, và không tiếp xúc (cảm biến siêu âm, radar...).

3. Xử lý ảnh dựa trên kỹ thuật học sâu

Như đã thảo luận ở trên, admin sẽ là người ra quyết định điều khiển mobile robot thông qua cử chỉ tay. Do đó, hệ điều khiển phải xử lý ba tác vụ chính như sau. Đầu tiên là tác vụ nhận diện khuôn mặt để nhận dạng quản trị viên. Thứ hai là phát hiện và xác định vị trí của quản trị viên so với hệ tọa độ gắn trên khung mobile robot, từ đó cung cấp thông tin cho bộ điều khiển cấp dưới trong chế độ điều khiển xe bám theo người. Nhiệm vụ cuối cùng là phát hiện tư thế tay và chuyển thành các mệnh lệnh di chuyển xe theo ý muốn người điều khiển.



Hình 3. Mô hình phát hiện và nhận dạng khuôn mặt dựa trên VGG-16

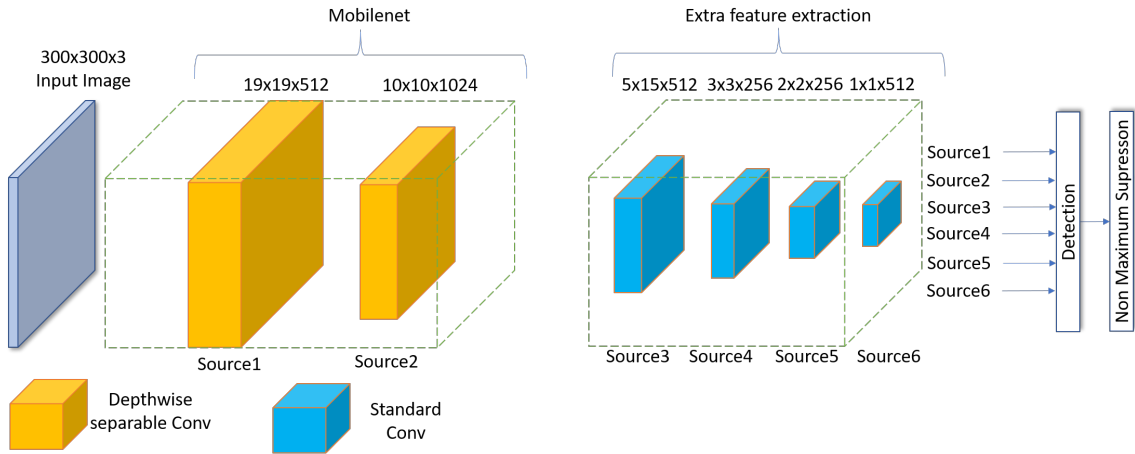
3.1. Nhận dạng khuôn mặt bằng mạng neural tích chập

Mạng nơ-ron tích chập (CNN) là một trong những mô hình học sâu thành công nhất dùng để nhận dạng khuôn mặt. Điểm khác biệt chính giữa CNN và mạng neural kinh điển là ở các lớp tích chập. Có thể coi đây là một bộ lọc vừa giúp trích xuất các đặc trưng của hình ảnh đầu vào, vừa giảm số lượng tham số cần xử lý. Trong ứng dụng này, mạng CNN với kiến trúc VGG-16 do nhóm nghiên cứu tại đại học Oxford phát triển được sử dụng [9]. Đầu vào của mạng là một hình ảnh có kích thước (224, 224, 3). Việc sử dụng các lớp tích chập với nhiều bộ lọc 3×3 , kèm sau đó là các lớp gộp 2×2 kiểu max pooling giúp giảm kích thước các lớp tiếp theo đi một nửa mà vẫn bảo toàn được các đặc trưng của hình ảnh. Cuối cùng, các đặc trưng đã trích xuất được đưa qua các lớp duỗi thẳng (biến ma trận thành vector) và đưa vào các lớp kết nối đầy đủ (FC) của mạng neural kinh điển với lớp đầu ra cuối cùng có 1000 nơ ron. Mô hình này đạt vị trí số một về phát hiện đối tượng và vị trí số hai phân loại ảnh trong cuộc thi ILSVRC 2014. Kiến trúc đầu vào của mạng CNN được mô tả chi tiết ở hình 3, trong khi các lớp FC ở phía đầu ra của VGG-16 được thay đổi để có thể vừa phát hiện ra khuôn mặt, vừa cung cấp tọa độ ô vuông chứa khuôn mặt trong hình ảnh đầu vào. Trong đó, lớp duỗi thẳng đầu ra từ VGG-16 được chia làm 2 nhánh. Một nhánh kết nối với lớp FC có 2048 nơ ron và một hàm softmax đầu ra để xác định xác suất p của khuôn mặt, nhánh còn lại gồm 2 lớp FC có kích thước lần lượt là 2048 và 4 tương ứng tọa độ $r = [x_1, x_2, y_1, y_2]$ của ô vuông chứa khuôn mặt.

Với dữ liệu đầu vào là 888 bức hình tự chụp và dán nhãn, kết quả huấn luyện theo phương pháp gradient decent của mô hình cho kết quả khả quan với độ chính xác tới 0.95. Sau khi có được vị trí của khuôn mặt, phương pháp phân tích thành phần chính (Principal Component Analysis - PCA) [10, 11] được sử dụng để nhận diện khuôn mặt tại vị trí đã được xác định xem có phải người đó là admin không. Với đặc thù hệ điều khiển dành cho mobile robot thường đòi hỏi chi phí thấp, dẫn đến năng lực tính toán hạn chế, phương pháp PCA đặc biệt phù hợp do kỹ thuật tính toán đơn giản cũng như dữ liệu của admin thường hạn chế ở một vài người [12].

3.2. Phát hiện và định vị người bằng mạng Mobilenet-SSD

Phát hiện con người trong một bức ảnh bao gồm hai tác vụ, phân loại (classification) và định vị (localization) bằng cách vẽ ô vuông bao quanh vị trí hình ảnh được xác định là "con người". Với các hệ thống có năng lực tính toán giới hạn, các thuật toán dựa trên kỹ thuật học sâu như Faster R-CNN [13], Single Shot Detectors (SSD) [14], và You Only Look Once (YOLO) [15] chiếm ưu thế nhờ khả năng xử lý tính toán nhanh [16]. Để thỏa mãn yêu cầu về khả năng xử lý thời gian thực trong khi vẫn đạt độ chính xác tương đối cao, Mobilenet SDD được lựa chọn do mô hình này có khả năng phân loại được nhiều đối tượng với tỉ lệ kích thước khác nhau trên một khung hình. Kiến trúc của mạng Mobilenet SDD được thể hiện ở hình 4. Khác với mạng VGG-16, Mobilenet sử dụng các lớp tích chập tách biệt chiều sâu (depthwise separable convolution layers) ở phía đầu vào, tức là mỗi kênh dữ liệu đầu vào sẽ áp dụng một bộ lọc khác nhau và hoàn toàn không chia sẻ tham số. Do đó giúp quá trình học và nhận diện đặc trưng sẽ được tách biệt theo từng bộ lọc. Khi đó, nếu đặc trưng trên các kênh là khác xa nhau thì sử dụng các bộ lọc riêng sẽ mang lại hiệu quả cao hơn trong việc phát hiện các đặc trưng. Các lớp tích chập thông thường ở đầu



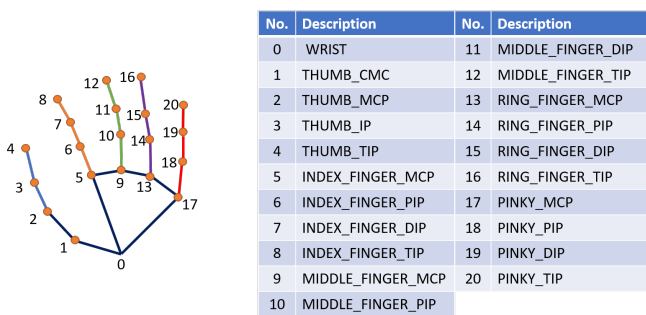
Hình 4. Kiến trúc mạng Mobilenet-SSD

ra (extra feature extraction layers) tạo thêm dữ liệu hỗ trợ ra việc phát hiện các đặc trưng trong bức ảnh. Đầu ra cuối cùng của Mobilenet SDD là một vector $y^T = [x, y, w, h, p]$ chứa tọa độ x, y , kích thước dài rộng w, h và xác suất của người được phát hiện trong bức ảnh.

Cuối cùng, tọa độ tâm của ô vuông bao quanh hình ảnh người sẽ được sử dụng để tính toán khoảng cách và hướng tới mobile robot dựa trên camera có cảm biến chiều sâu D435i. Thông tin về hướng và khoảng cách được sử dụng bởi bộ điều khiển xe nằm ở cấp thấp hơn.

3.3. Nhận diện tư thế tay

Để giải quyết bài toán hỗ trợ con người, ngoài việc bám theo admin thì việc di chuyển theo ý muốn bằng cách nhận diện cử chỉ tay cũng đóng vai trò quan trọng, đặc biệt là ở môi trường không gian hẹp và nằm ngoài phạm vi bản đồ đã cài đặt sẵn trên xe. Do đó nhóm nghiên cứu tận dụng module MediaPipe Hand [20] đã được huấn luyện sẵn để phát hiện các điểm đặc biệt của bàn tay như mô tả ở hình 5, từ đó hình thành nên các lệnh di chuyển như "chạy", "dừng", "rẽ trái", "rẽ phải", "tiến", "lùi", "quay tròn"...



Hình 5. Mô hình khung xương của bàn tay

4. Thiết kế hệ điều khiển

4.1. Điều khiển hướng di chuyển của xe

Với việc mỗi bánh xe được điều khiển bởi một hệ truyền động servo, đồng thời giả thiết là các bánh xe tiếp xúc tốt với mặt sàn sao cho không có hiện tượng trượt, khi đó ta có thể giả thiết

rằng mối quan hệ giữa vận tốc mong muốn và vận tốc thực của xe là một khâu quán tính bậc nhất như (7)

$$\begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix} = \left(\frac{1}{1 + T_{drv}s} \right) \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_1^* \\ \theta_2^* \\ \theta_3^* \\ \theta_4^* \end{bmatrix} \quad (7)$$

trong đó T_{drv} là hằng số thời gian của bộ biến đổi.

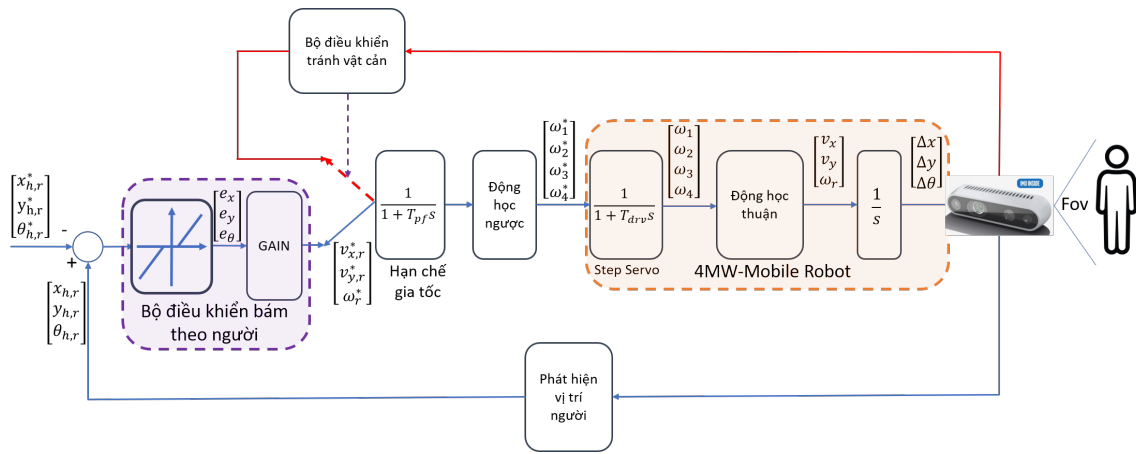
Dựa trên (2) và (7), mối quan hệ giữa quãng đường dịch chuyển của xe và tốc độ góc của từng bánh xe có thể được mô tả bởi (8)

$$\begin{bmatrix} \Delta x_r \\ \Delta y_r \\ \Delta \Phi_r \end{bmatrix} = \frac{1}{s} \frac{1}{(1 + T_{drv}s)} R J^+ \mathbf{I} \begin{bmatrix} \theta_1^* \\ \theta_2^* \\ \theta_3^* \\ \theta_4^* \end{bmatrix} \quad (8)$$

Ở chế độ bám theo admin, mục tiêu điều khiển là giữ khoảng cách an toàn giữa người và mobile robot, cụ thể hơn là ta cần bám lượng đặt $q_{h,r}^* = [x_{h,r}^*, y_{h,r}^*, \theta_{h,r}^*]^T$ trong khung tọa độ thân xe như mô tả ở hình 1. Để thấy với sự xuất hiện của một khâu tích phân trong (8), một bộ điều khiển tỉ lệ là đủ để đảm bảo mục tiêu điều khiển trên. Do có ràng buộc về vận tốc cực đại của xe, hệ số tỉ lệ K_P nên được hiệu chỉnh trong khoảng:

$$\begin{aligned} 0 < e_{x,max} K_P x &\leq v_{xr,max} \\ 0 < e_{y,max} K_P y &\leq v_{yr,max} \\ 0 < e_{\theta,max} K_P \theta &\leq \omega_{r,max} \end{aligned} \quad (9)$$

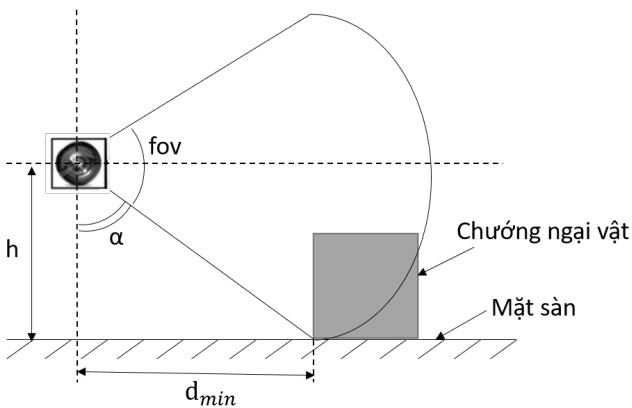
trong đó, $e_{x,max}$, $e_{y,max}$ và $e_{\theta,max}$ lần lượt là sai lệch cực đại theo các trục. Và $v_{xr,max}$, $v_{yr,max}$, $\omega_{r,max}$ lần lượt là giới hạn vận tốc theo các trục của xe. Bất đẳng thức (9) có thể được hiểu một cách rất đơn giản là xe sẽ đạt tới vận tốc cực đại nếu sai lệch bám lớn hơn ngưỡng cho phép. Trong thực tế, nếu admin đứng tại chỗ và khi xe đạt tới khoảng cách an toàn mong muốn, hiện tượng rung có thể xảy ra do dữ liệu khoảng cách lấy về từ camera không thực sự ổn định. Do đó một khâu "dead-band" được thêm vào để đảm bảo xe dừng hoàn toàn khi sai lệch vị trí đạt tới giá trị "chấp nhận được". Đồng thời một khâu quán tính cũng được thêm vào ngay sau khâu khuếch đại để hạn chế gia tốc của xe, tránh hiện tượng giật mạnh khi admin chuyển từ trạng thái đứng im sang di chuyển. Do đó, cấu trúc hệ điều khiển điều hướng xe bám theo người được đề xuất như mô tả ở hình 6.



Hình 6. Hệ điều khiển xe bám theo người và tránh vật cản

4.2. Điều khiển tránh vật cản

Trong quá trình bám theo người sử dụng, tình huống gặp vật cản trên đường đi thường xuyên xảy ra. Do đó nhóm nghiên cứu đề xuất một thuật toán đơn giản để tránh vật cản dựa trên xử lý ảnh cũng như khả năng di chuyển đa hướng của 4-MWMR. Giả



Hình 7. Vị trí đặt camera so với vật cản

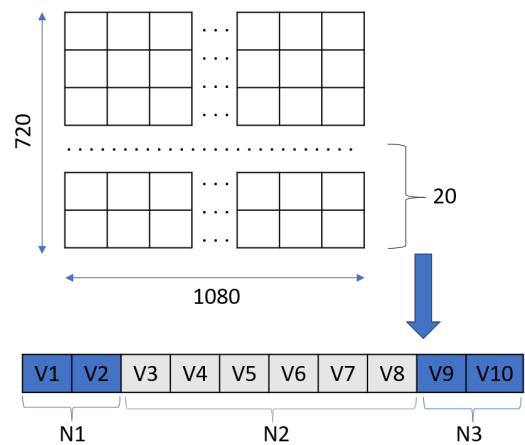
sử mặt sàn là bằng phẳng và vị trí đặt camera so với chướng ngại vật như thể hiện ở hình 7. Khi đó, khoảng cách tối thiểu d_{min} để camera phát hiện chướng ngại vật được xác định bởi phương trình:

$$\alpha = 90^\circ - \frac{fov}{2} \tag{10}$$

$$d_{min} = h \tan \alpha$$

Trong đó, h là chiều cao đặt camera và $fov = 58^\circ$ là trường nhìn dọc của camera.

Dữ liệu thu được từ D435i là một ma trận chứa thông tin khoảng cách đến các điểm ảnh tương ứng, có kích thước 1080×720 . Khoảng cách được biểu diễn bằng các số nguyên có đơn vị là milimet. Để đơn giản, nghiên cứu này chỉ xét các vật cản tính từ mặt đất lên một độ cao nhất định tương đương chiều cao của xe. Do đó dữ liệu được xét đến ở đây là hai mươi hàng dưới cùng của ma trận điểm ảnh. Ma trận kích thước 1080×20 được chia thành mười vùng ($V_i, i = 1 \dots 10$) như hình 8, mỗi vùng có kích thước 108×20 . Mỗi phần tử của từng vùng được so sánh với d_{min} , nếu có tối thiểu 10 phần tử nhỏ hơn d_{min} thì vùng đó được định nghĩa là “có chướng ngại vật”, tương ứng



Hình 8. Phân vùng dữ liệu để phát hiện vật cản

giá trị $V_i = 1$. Mặt khác, nếu có ít hơn 10 hoặc không có phần tử nào nhỏ hơn d_{min} thì vùng đó được coi là “không chướng ngại vật”, tương ứng với $V_i = 0$. Tiếp theo, các vùng được gộp lại thành ba nhóm ($N_i, i = 1 \dots 3$) như mô tả ở hình 8, trong đó:

$$N_1 = V_1 \cup V_2$$

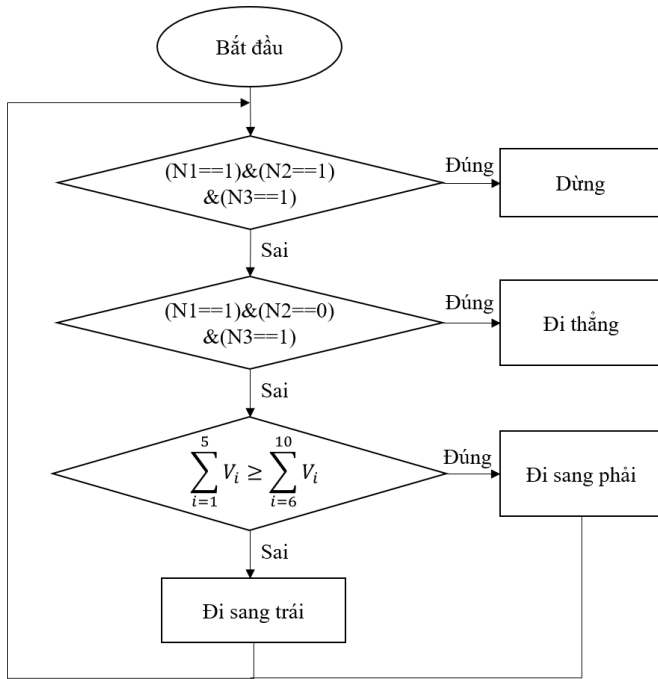
$$N_2 = V_3 \cup V_4 \cup V_5 \cup V_6 \cup V_7 \cup V_8 \tag{11}$$

$$N_3 = V_9 \cup V_{10}$$

Dựa vào các thông tin thu được từ (11), hệ điều khiển sẽ đưa ra lựa chọn cho việc xe tiếp tục di chuyển thẳng, hay cần rẽ sang một trong hai bên để tránh vật cản theo lưu đồ thể hiện ở hình 9. Trong quá trình tránh vật cản, lượng đặt tốc độ cho mobile robot được tính theo bảng 2, trong đó có thể hiểu là đối tượng cần bám càng xa xe thì càng cần tránh nhanh chướng ngại vật để đuổi theo.

Bảng 2. Bảng tính lượng đặt tốc độ xe ở chế độ tránh vật cản

Trạng thái	Lượng đặt tốc độ
Rẽ phải	$v_x^* = K_{Px} \sqrt{e_x^2 + e_y^2}, v_y^* = \omega^* = 0$
Rẽ trái	$v_x^* = -K_{Px} \sqrt{e_x^2 + e_y^2}, v_y^* = \omega^* = 0$
Dừng	$v_x^* = v_y^* = \omega^* = 0$



Hình 9. Lưu đồ thuật toán tránh vật cản

4.3. Điều khiển logic vận hành

Cuối cùng, để đảm bảo các thuật toán điều khiển mobile robot vận hành phối hợp với nhau, hệ điều khiển logic được thiết kế dựa trên kỹ thuật mô tả trạng thái máy (state-machine) như mô tả ở hình 10. Khi mới được cấp nguồn, xe ở trạng thái



Hình 10. Điều khiển phối hợp các chế độ vận hành xe

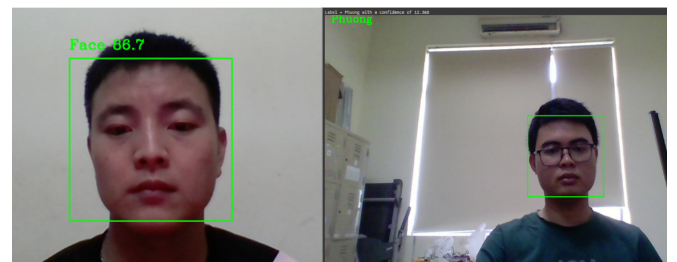
dừng. Nếu có lệnh *chạy*, xe chuyển sang trạng thái "đò tìm khuôn mặt". Nếu phát hiện ra người được cấp quyền (admin), xe sẽ căn cứ vào hiệu lệnh tay của "admin" để chọn chế độ điều khiển "bằng tay" hay "bấm người". Trong quá trình "bấm người", việc tính chọn lượng đặt tốc độ cho từng bánh xe sẽ do bộ điều khiển thiết kế ở 4.1 nếu không có vật cản giữa xe và người. Ngược lại, tốc độ đặt sẽ được tính bởi bảng 2.

5. Các kết quả thực nghiệm

Dựa trên nguyên mẫu xe như minh họa ở hình 11, nhóm đã tiến hành một số thực nghiệm. Để phát hiện khuôn mặt, mô hình cho thấy kết quả khá tốt khi tổn thất trung bình trên toàn bộ tập dữ liệu đào tạo với 888 hình ảnh nhỏ hơn 0.1, và độ chính xác đạt được lớn hơn 0.95. Kết quả phát hiện và nhận diện admin được minh họa ở hình 12, trong đó phép đo độ tin cậy là 86.7 và vị trí của hộp giới hạn là chính xác.



Hình 11. Hệ thống thực nghiệm



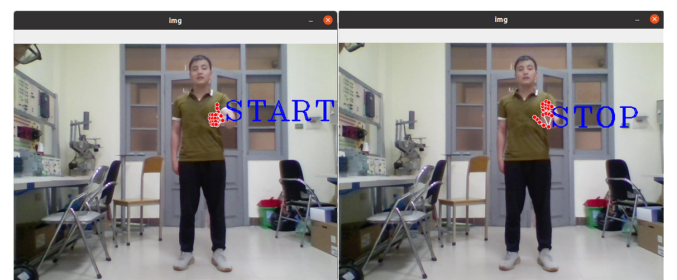
(a) Face detection (b) Face recognition

Hình 12. Phát hiện và nhận diện khuôn mặt

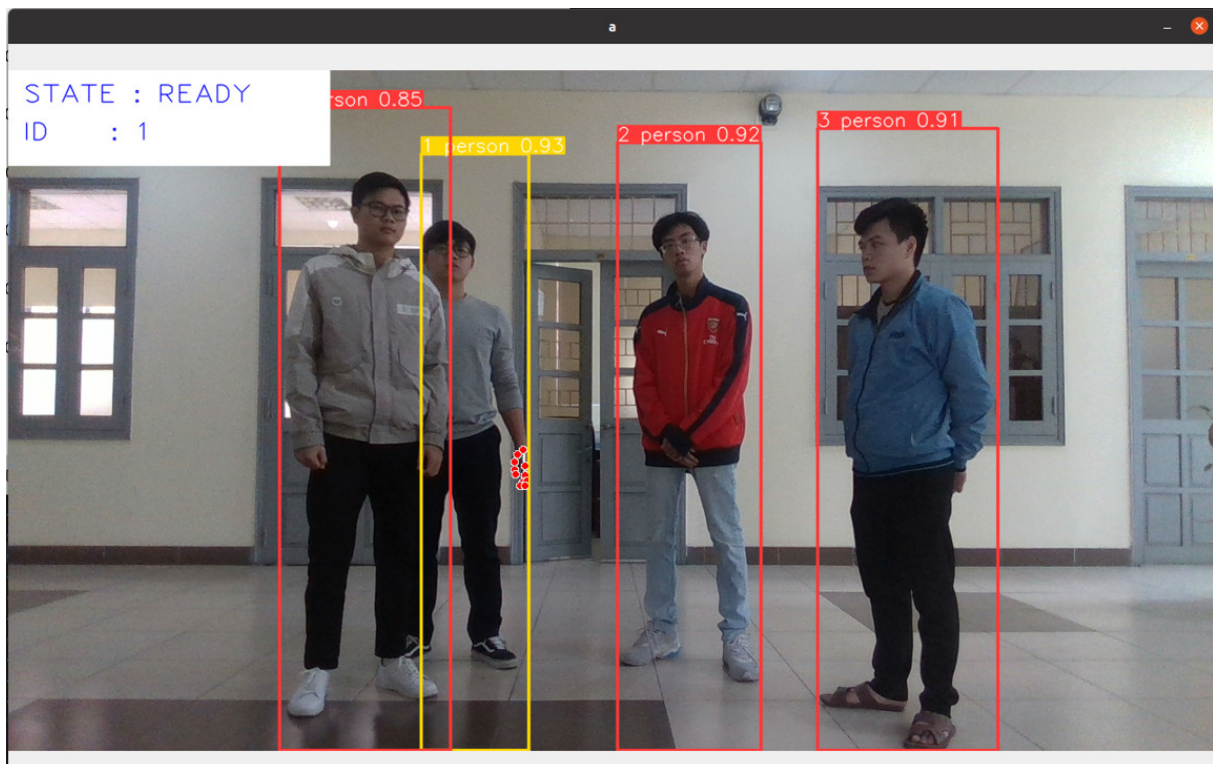
Hình 13 và 14 cho thấy kết quả phát hiện con người cũng như tư thế bàn tay. Có thể thấy admin có thể được nhận dạng ngay cả khi một phần cơ thể bị che khuất. Dựa trên kết quả dự đoán tốt, robot di động được thử nghiệm hoạt động ở cả chế độ điều khiển bấm theo người và điều khiển tư thế tay. Sau khi đã được nhận dạng khuôn mặt và hình dáng, admin có thể quay lưng lại mà xe vẫn có thể bám theo như mô tả ở hình 15. Kết quả thử nghiệm với mô hình xe có thể được xem chi tiết theo đường link: <https://www.youtube.com/watch?v=PPirzwVJ3WI>

Ở chế độ tránh vật cản, đầu tiên dữ liệu phân vùng của vật cản được kiểm tra như mô tả ở hình 16. Việc đặt vật cản ở các vị trí biết trước giúp hiệu chỉnh cảm biến hình ảnh cũng như thuật toán trước khi bắt đầu vận hành. Có thể thấy các kết quả thu được cho thấy thuật toán hoạt động tốt khi xác định được vị trí vật cản.

Sau khi đã hiệu chỉnh xong cảm biến, nhóm nghiên cứu tiến hành thử nghiệm chế độ tránh vật cản theo kịch bản mô tả ở hình 17, trong đó xe sẽ thực hiện



Hình 13. Nhận diện lệnh điều khiển qua tư thế tay



Hình 14. Phát hiện người trong khung hình



Hình 15. Quá trình xe bám theo người

việc bám theo người đồng thời lách qua một và hai vật cản. Video thực nghiệm được cung cấp theo đường link https://drive.google.com/file/d/17aLHquEv-WvwUkH9W2KN6m4gm00nXu5/view?usp=share_link

Các dữ liệu thu về từ quá trình thực nghiệm được thể hiện ở hình 17 cho thấy xe vẫn duy trì bám theo người và dừng lại khi đạt tới khoảng cách an toàn được cài đặt từ trước. Mặc dù trong quá trình tránh vật cản, khoảng cách tới người bị dẫn ra xa do xe ưu tiên tác vụ "tránh vật cản" hơn.

6. Kết luận

Trong nghiên cứu này, một robot di động đa hướng nhằm hỗ trợ con người được phát triển. Bằng cách sử dụng các kỹ thuật xử lý hình ảnh dựa trên kỹ thuật học sâu như VGG-16, mobilenet-SSD, kèm theo một thuật toán tránh vật cản do nhóm nghiên cứu đề xuất, mô hình xe đã có thể thực hiện các chức năng chính như nhận dạng, bám theo quản trị viên, tránh chướng

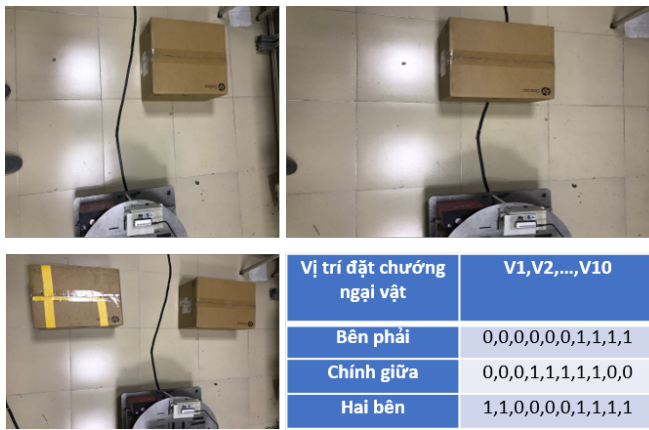
ngại vật và điều khiển dựa trên tư thế tay. Các kết quả thực nghiệm đạt được cho thấy tiềm năng ứng dụng của robot di động trong thực tế.

Lời cảm ơn

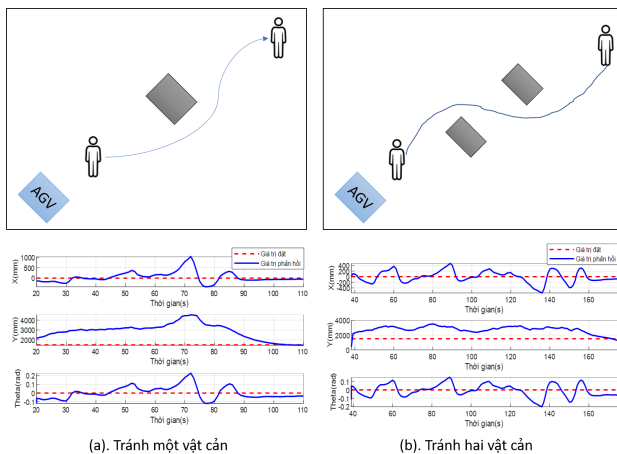
Nghiên cứu này được hỗ trợ bởi Đại học Bách khoa Hà Nội, thuộc đề tài mã số T2022-PC-005.

Tài liệu tham khảo

- [1] Rubio F, Valero F, Llopis-Albert C. A review of mobile robots: Concepts, methods, theoretical framework, and applications. *International Journal of Advanced Robotic Systems*. 2019;16(2). doi:10.1177/1729881419839596
- [2] Hamid Taheri, Chun Xia Zhao, "Omnidirectional mobile robots, mechanisms and navigation approaches," *Mechanism and Machine Theory*, Volume 153, 2020, 103958, ISSN 0094-114X, <https://doi.org/10.1016/j.mechmachtheory.2020.103958>.
- [3] Sergio Cebollada, Luis Payá, María Flores, Adrián Peidro, Oscar Reinoso, "A state-of-the-art review on mobile robotics tasks



Hình 16. Kiểm tra dữ liệu phân vùng vật cản



(a). Tránh một vật cản

(b). Tránh hai vật cản

Hình 17. Kịch bản thử nghiệm tránh vật cản

using artificial intelligence and visual data,” *Expert Systems with Applications*, Volume 167, 2021, 114195, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2020.114195>.

- [4] L. Payá, A. Gil, O. Reinoso, "A State-of-the-Art Review on Mapping and Localization of Mobile Robots Using Omnidirectional Vision Sensors", *Journal of Sensors*, vol. 2017, Article ID 3497650, 20 pages, 2017. <https://doi.org/10.1155/2017/3497650>.
- [5] Prabin Kumar Panigrahi, Sukant Kishoro Bisoy, "Localization strategies for autonomous mobile robots: A review," *Journal of King Saud University - Computer and Information Sciences*, Volume 34, Issue 8, Part B, 2022, Pages 6019-6039, ISSN 1319-1578, <https://doi.org/10.1016/j.jksuci.2021.02.015>.
- [6] M. Gupta, S. Kumar, L. Behera and V. K. Subramanian, "A Novel Vision-Based Tracking Algorithm for a Human-Following Mobile Robot," in *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 7, pp. 1415-1427, July 2017, doi: 10.1109/TSMC.2016.2616343.
- [7] D. Jin, Z. Fang and J. Zeng, "A Robust Autonomous Following Method for Mobile Robots in Dynamic Environments," in *IEEE Access*, vol. 8, pp. 150311-150325, 2020, doi: 10.1109/ACCESS.2020.3016472.
- [8] Yuan, Z., Tian, Y., Yin, Y., Wang, S., Liu, J. and Wu, L. (2020), Trajectory tracking control of a four mecanum wheeled mobile platform: an extended state observer-based sliding mode approach. *IET Control Theory Appl.*, 14: 415-426. <https://doi.org/10.1049/iet-cta.2018.6127>
- [9] <https://www.geeksforgeeks.org/vgg-16-cnn-model/>
- [10] Erwin et al, "A Study about Principle Component Analysis and Eigenface for Facial Extraction", 2019 *J. Phys.: Conf. Ser.* 1196 012010, doi: 10.1088/1742-6596/1196/1/012010
- [11] KIM, Kyungnam. Face recognition using principle component analysis. In: *International Conference on Computer Vision and Pattern Recognition*. 1996. p. 591
- [12] Schenkel, O. Ringhage, and N. Branding, "A Comparative Study of Facial Recognition Techniques: With focus on low computational power," *Dissertation*, 2019
- [13] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, 2017.

- [14] W. Liu, D. Anguelov, D. Erhan et al., "SSD: single shot multibox detector," in *European Conference on Computer Vision*, pp. 13-17, Springer, 2016.
- [15] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779-788, Las Vegas, USA, 2016.
- [16] L. I. Yinan, "A survey of research on deep learning target detection methods," *China New Telecommunications*, vol. 23, no. 9, pp. 159-160, 2021.
- [17] Y. -C. Chiu, C. -Y. Tsai, M. -D. Ruan, G. -Y. Shen and T. -T. Lee, "Mobilenet-SSDv2: An Improved Object Detection Model for Embedded Systems," *2020 International Conference on System Science and Engineering (ICSSE)*, 2020, pp. 1-5, doi: 10.1109/IC-SSE50014.2020.9219319.
- [18] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications." [Online]. Available: <http://arxiv.org/abs/1704.04861>
- [19] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks." [Online]. Available: <http://arxiv.org/abs/1801.04381>
- [20] Zhang, Fan; Bazarevsky, Valentin; Vakunov, Andrey; Tkachenka, Andrei; Sung, George; Chang, Chuo-Ling and Grundmann, Matthias. (2020). *MediaPipe Hands: On-device Real-time Hand Tracking*.