NGHIÊN CỨU, THIẾT KẾ HỆ THỐNG AN NINH NHÀ THÔNG MINH ỨNG DỤNG CÔNG NGHỆ NHẬN DẠNG KHUÔN MẶT

RESEARCH AND DESIGN OF A SMART HOME SECURITY SYSTEM APPLYING FACIAL RECOGNITION TECHNOLOGY

Tạ Trọng Khánh[1], Đặng Thái Việt[1,*], Dương Thái Hoàng[1], Vi Trung Kiên[1]

DOI: https://doi.org/10.57001/huih5804.2024.033

TÓM TẮT

Trí tuệ nhân tạo và Internet kết nối vạn vật thu hút nhiều sự quan tâm của những học giả và nhà nghiên cứu, không chỉ bởi tính ứng dụng cao mà còn là những công nghệ tiêu biểu của Cách mạng công nghiệp lần thứ tư. Điểm nổi bật của trí tuệ nhân tạo là khả năng tự học, cho phép máy tính dự đoán và phân tích dữ liệu phức tạp như dấu vân tay, mống mắt và khuôn mặt. Nghiên cứu này đề xuất giải pháp hệ thống cơ điện tử tích hợp khả năng nhận dạng khuôn mặt của AI để tạo ra hệ thống an ninh nhà thông minh. Kết quả của nghiên cứu được so sánh với nghiên cứu gần đây chứng minh khả năng sử dụng của hệ thống. Vì vậy, nhóm tác giả vận dụng kết quả của quá trình đào tạo để xây dựng hệ thống an ninh nhà thông minh ứng dụng công nghệ nhận dạng khuôn mặt sử dụng mạng học sâu tích chập nơ ron.

Từ khóa: *Hệ thống an ninh nhà thông minh; internet kết nối vạn vật; nhận diện khuôn mặt; trí tuệ nhân tạo.*

ABSTRACT

Artificial intelligence and IoT have attracted a lot of attention from scholars and researchers, not only because of their high applicability but also because they represent key technologies of the Fourth Industrial Revolution. The notable feature of artificial intelligence is its ability to learn autonomously, enabling computers to predict and analyze complex data such as fingerprints, irises, and faces. This research proposes an integrated electromechanical system solution with AI facial recognition capabilities to create a smart home security system. The research results are compared with recent studies to demonstrate the system's effectiveness. Therefore, the authors apply the training results to build a smart home security system using facial recognition technology and convolutional neural networks.

Keywords: *Artificial intelligence; facial recognition; home security system; Internet of Things (IoT).*

[1]Đại học Bách khoa Hà Nội

[*]Email: viet.dangthai@hust.edu.vn

Ngày nhận bài: 05/6/2023

Ngày nhận bài sửa sau phản biện: 20/9/2023

Ngày chấp nhận đăng: 20/01/2024

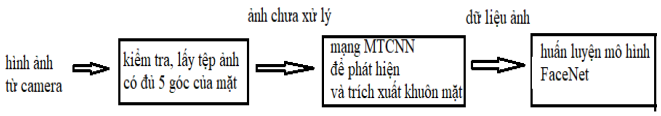
1. GIỚI THIỆU

Trong kỷ nguyên Công nghiệp 4.0, với sự phát triển nhanh chóng của các thiết bị phần cứng cung cấp khả năng tính toán cho các mạng trí tuệ nhân tạo. Các thành tựu khoa học kỹ thuật đã tạo sự phát triển của các kỹ thuật cơ bản trong trí tuệ nhân tạo. Trí tuệ nhân tạo với ứng dụng rộng rãi vào trong các thiết bị, hệ thống thông minh [1]. Một trong các ứng dụng quan trọng của trí tuệ nhân tạo là vấn đề bảo mật và phát triển hệ thống nhà thông minh [2].

Hiện nay có rất nhiều phương pháp bảo mật trong hệ thống nhà thông minh, như sử dụng sử dụng sinh trắc vân tay là một phương pháp hiện đại, có hiệu quả cao, tuy nhiên chia phí đầu tư lại cao, hệ thống thẻ dựa trên RFID mang lại sự tiện lợi và chính xác, nhưng chúng dễ bị các hoạt động gian lận,… [2]. Do đó để đảm bảo độ chính xác và bảo mật cao, nhóm tác giả đề xuất sử dụng phương pháp nhận diện bằng khuôn mặt kết hợp hệ thống RFID trong hệ thống nhà thông minh, nhóm tác giả tham khảo một số mô hình phổ biến như ArcFace, DeepFace và FaceNet [3, 4] và đề xuất ứng dụng mạng FaceNet, tuy có độ chính xác không cao bằng mạng ArcFace hay DeepFace nhưng bù lại tốn ít tài nguyên tính toán hơn, phù hợp triển khai đơn lẻ trên các hệ thống nhúng IoT [5, 6].

Cảnh báo xâm nhập bằng phương pháp phát hiện có người là một trong đề tài có tính ứng dụng cao trong các lĩnh vực như bảo mật, phân loại vật thể. Các thuật toán mới của phát hiện đối tượng (object detection) như YOLO, SSD có tốc độ khá nhanh và độ chính xác cao nên giúp cho mô hình có thể thực hiện được các tác vụ theo thời gian thực, tốc độ nhanh và đảm bảo độ chính xác và chất lượng hệ thống [7-9]. Các mô hình cũng trở nên tối ưu hơn nên cho phép hoạt động trên các thiết bị IoT để cấu thành nên hệ thống cơ điện tử thông minh. Để tạo điều kiện lấy mẫu hiệu quả cũng như tăng tính đa dạng của mẫu, đồng thời áp dụng làm tính bảo mật chống ảnh giả mạo, nhóm đã tìm hiểu, kết hợp thêm với thuật toán xác định góc quay khuôn mặt từ nhiều góc nhìn, kết quả được gửi đến máy tính nhúng JetsonNano để xử lý ứng dụng nhà thông minh.

2. PHƯƠNG PHÁP NGHIÊN CỨU

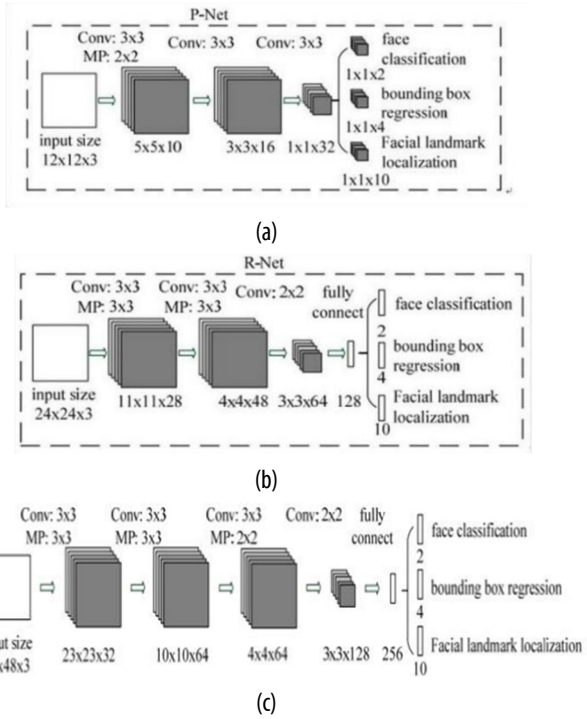


Hình 1. Phương pháp thu thập dữ liệu

Nhóm tác giả dùng phương pháp thu thập dữ liệu như hình 1.

2.1. Mô hình mạng tích chấp xếp tầng đa tác vụ (MTCNN)

MTCNN (Multi-tasked Cascaded Convolutional Networks) là một lựa chọn phổ biến để sử dụng trong Deep Convolutional Neural Networks (DCNN) để phát hiện khuôn mặt và nhận dạng đặc điểm khuôn mặt [6]. Mô hình MTCNN bao gồm ba giai đoạn liên tiếp: phát hiện khuôn mặt (a), định vị mốc khuôn mặt (b) và hồi quy hộp giới hạn chính xác (c). Các giai đoạn này được thiết kế để phát hiện và định vị các khuôn mặt trong hình ảnh với các mức độ phức tạp khác nhau, bao gồm các biến thể về tỷ lệ, tư thế và ánh sáng.



Hình 2. Cấu trúc mạng MTCNN bao gồm 3 lớp: phát hiện khuôn mặt (a), định vị mốc khuôn mặt (b), và hồi quy hộp giới hạn chính xác (c)

2.2. Mô hình mạng Facenet

FaceNet sử dụng mạng Nơ-rơn tích chập (CNN), là một hệ thống nhận dạng khuôn mặt tự động, có khả năng biểu diễn mỗi khuôn mặt thành một vectơ nhúng (embedding vector) nhiều chiều. Để đảm bảo những ảnh đầu vào của cùng một người sẽ cho những vector nhúng có khoảng cách đủ gần, Facenet sử dụng hàm Triplet loss. Mô hình mạng FaceNet học theo các bước sau thông qua hàm triplet loss, trong đó nó sử dụng 3 ảnh đầu vào bao gồm: Anchor face, Positive face, Negative face. Trong đó ảnh đầu vào cần vector hóa là Anchor face, ảnh của cùng một người với ảnh đầu vào là Positive face và Negative face là ảnh của một người khác với người trong ảnh đầu vào Anchor.

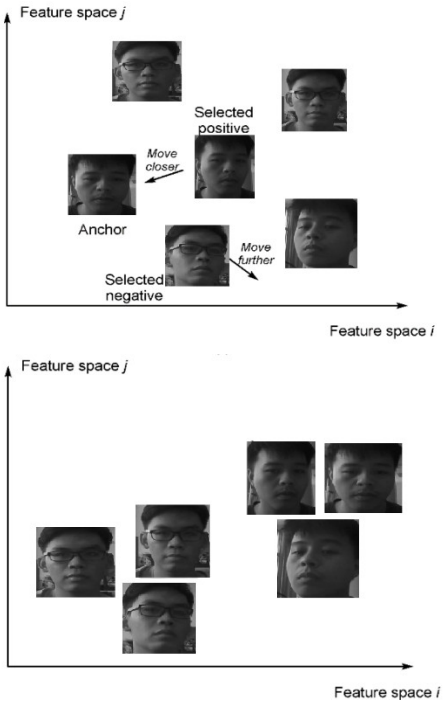
- Bước 1. Chọn ngẫu nhiên một ảnh làm ảnh đánh dấu (anchor image).

- Bước 2. Chọn ngẫu nhiên một hình ảnh của cùng một người làm hình ảnh đánh dấu làm hình ảnh tích cực (positive image).

- Bước 3. Chọn ngẫu nhiên một hình ảnh của một người khác với hình ảnh đánh dấu làm hình ảnh tiêu cực (negative image).

- Bước 4. Điều chỉnh các thông số của mạng Facenet sao cho hình ảnh tích cực gần với hình ảnh đánh dấu hơn hình ảnh tiêu cực.

Ví dụ: dùng tệp ảnh có nhiều vị trí khuôn mặt để training mạng FaceNet, hình ảnh từ 2 người khác nhau, với điều kiện đã xác định được ảnh là của ai, khi chọn ảnh ngẫu nhiên, ảnh tích cực sẽ được di chuyển lại gần ảnh đánh dấu và ngược lại với ảnh tiêu cực.



Hình 3. Mô tả quá trình huấn luyện FaceNet

$$\left\|f\left(x_i^a\right)-f\left(x_i^p\right)\right\|_2^2+\alpha<\left\|f\left(x_i^a\right)-f\left(x_i^n\right)\right\|_2^2, \quad \forall f\left(x_i^a\right), f\left(x_i^p\right), f\left(x_i^n\right)\in T \tag{1}$$

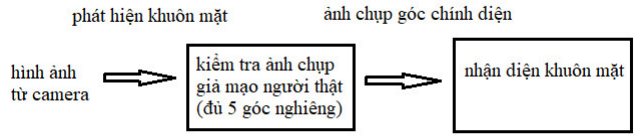
Trong đó tham số α được sử dụng để đảm bảo khoảng cách giữa các lần nhúng của các mẫu khác nhau là đủ xa. Điều này giúp tăng độ chính xác của hệ thống phân loại và giảm sự phụ thuộc vào các tính năng không cần thiết. Khi đó, hàm loss được biểu diễn như sau:

$$L=\sum_i^N\left[\left\|f\left(x_i^a\right)-f\left(x_i^p\right)\right\|_2^2-\left\|f\left(x_i^a\right)-f\left(x_i^n\right)\right\|_2^2\right]+\alpha \tag{2}$$

Với các tập dữ liệu lớn, việc chọn các điểm tích cực và tiêu cực thích hợp từ toàn bộ tập dữ liệu có thể tốn kém về mặt tính toán, các tác giả chia tập dữ liệu thành các bộ dữ liệu nhỏ và chọn các điểm tích cực và tiêu cực phù hợp.

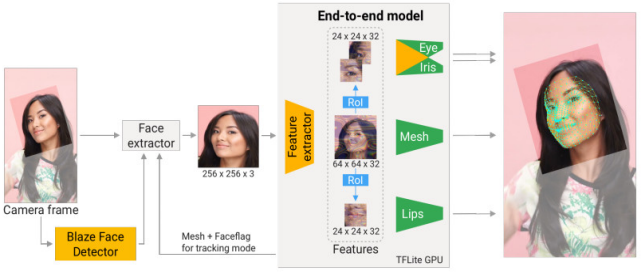
2.3. Các mô hình phát hiện, nhận diện khuôn mặt và tính toán góc nghiêng

Nhóm tác giả kết hợp sử dụng phương pháp tính toán góc nghiêng khuôn mặt để tăng cường thêm cho phòng chống vấn đề giả mạo người thật bằng ảnh chụp.



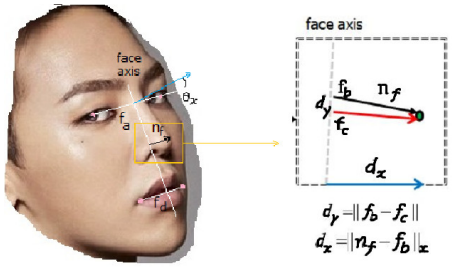
Hình 4. Mô tả quá trình kiểm tra giả mạo và nhận diện khuôn mặt

Attention mesh là một mạng neural dùng để xác định vị trí của khuôn mặt bằng cách sử dụng những khu vực đặc thù trên khuôn mặt qua các phép biến đổi không gian.

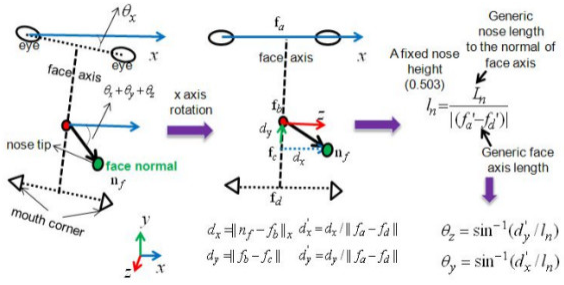


Hình 5. Cấu trúc mạng Attention mesh.

Đầu vào từ camera sẽ được xử lý để tìm ra bức ảnh có chứa khuôn mặt với kích thước 256x256x3, sau đó model này sẽ chia ra làm những model nhỏ hơn để tìm những đặc trưng của khuôn mặt. Đặc biệt kết quả tập trung vào 2 vùng là môi và 2 mắt từ đó đoán được 478 điểm trong khuôn mặt 3D và xác định vùng đặc trưng khác của khuôn mặt.



(a)



(b)

Hình 6. Cấu trúc tập dữ liệu khuôn mặt 3D và mô phỏng cách tính toán góc nghiêng khuôn mặt (a), (b) Mô phỏng cách tính toán góc nghiêng khuôn mặt

Sử dụng phương pháp tính toán dựa trên phương pháp hình học trong đó lấy 5 điểm là 2 mắt, mũi, 2 khoé miệng, cùng với các khoảng cách được chiếu xuống của khuôn mặt làm đặc trưng. Đường bình thường của khuôn mặt n_f được xác định bằng cách nối 2 điểm chính giữa của đường nối 2 mắt và miệng lại với nhau, dùng một điểm xác định trước trên đường đó để nối với đầu mũi. Để tính toán các giá trị như roll, pitch, yaw của khuôn mặt, ta cần xác định 4 điểm (f_a, f_b, f_c, f_d). Với f_a, f_d là hai điểm chính giữa của mắt và khoé miệng, đường giữa f_a và f_d là trục của khuôn mặt. f_b là điểm trực giao của của trục khuôn mặt với pháp tuyến n_f của gương mặt. f_c được tính toán bằng cách:

$f_c = (1-t) * f_a + t * f_d$ trong đó t được chọn dựa trên tỉ lệ

$$n_f = \frac{\|f_b - f_d\|}{\|f_a - f_d\|} \sim 0,45 \quad (3)$$

Hoặc tỉ lệ: $\frac{f_c f_d}{f_a f_d} = 0,4 (R_m = \frac{L_m}{L_f} = 0,4)$. Khoảng cách chiếu xuống của đường pháp tuyến n_f trong trục x và y:

$$d_y = \|f_b - f_c\|, d_x = \|n_f - f_b\|_x \quad (4)$$

Dựa vào dữ liệu của mô hình 3D, khoảng cách phía trên được chuẩn hoá dựa trên chiều cao của mũi và khoảng cách dự kiến của yaw, pitch khi n_f chiếu xuống mặt phẳng:

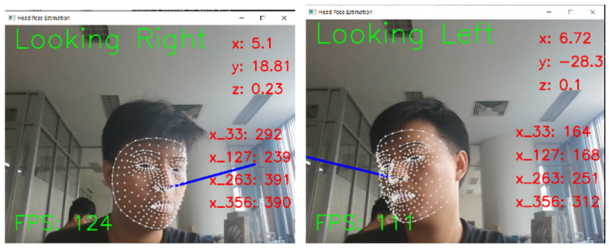
$$d_y' = \frac{\|f_b - f_c\|}{\|f_a - f_d\|}, d_x' = \frac{\|n_f - f_b\|_x}{\|f_a - f_d\|} \quad (5)$$

Chiều cao của mũi: $l_n = \frac{L_n}{\|f_a' - f_d'\|} \sim 0,5$ (6)

với f_a', f_d' có được từ mô hình 3D trung bình, L_n là chiều dài được chuẩn hoá dựa vào chiều dài khuôn mặt. Cuối cùng tính toán được góc yaw và pitch:

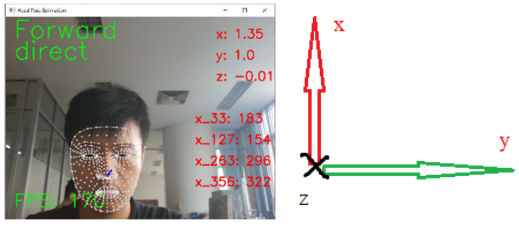
$$\theta_y = \arcsin\left(\frac{d_x'}{l_n}\right), \theta_z = \arcsin\left(\frac{d_y'}{l_n}\right) \quad (7)$$

Khi muốn truy cập vào hệ thống, người đó phải cung cấp đủ 5 góc của khuôn mặt. Tính toán dựa vào góc nghiêng của khuôn mặt với trục y là trục theo chiều ngang và x là trục theo chiều dọc. Các góc được tính bằng cách so với trục z là trục vuông góc với màn hình (tương ứng lấy từ các góc yaw và pitch).



(a) (b)

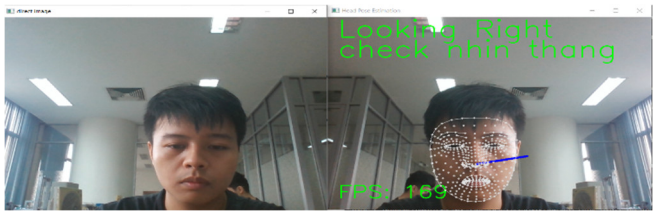
(c) (d)



(e) (f)

Hình 7. góc nghiêng phải (10 độ) (a), góc nghiêng trái (b) (-10 độ), góc nhìn xuống (c) (-10 độ), góc nhìn lên (d) (10 độ), góc nhìn thẳng (e), trục toạ độ (f)

Sau khi kiểm tra lần lượt các góc, góc nhìn thẳng sẽ được trọn để nhận diện khuôn mặt như hình 8.

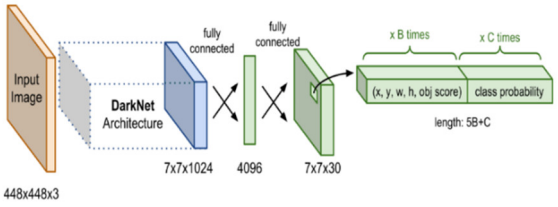


Hình 8. Kiểm tra tính toán góc nghiêng khuôn mặt

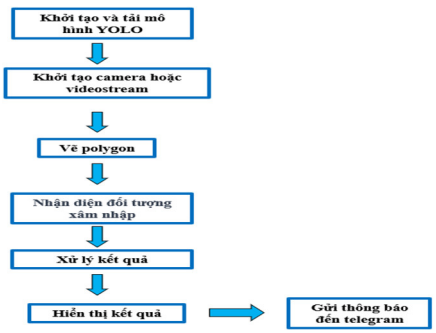
Hình ảnh đầu vào và các phương trình (4) - (8) sẽ phát hiện ra khuôn mặt, biểu diễn các góc quay của khuôn mặt người theo các tư thế đầu đặc trưng. Kết quả tăng cường cho kỹ thuật nhận dạng khuôn mặt, tăng độ chính xác và chống ảnh giả mạo người thật.

2.4. Cảnh báo xâm nhập sử dụng YOLOv3

YOLO (You Only Look Once) là một thuật toán object detection tiên tiến. Kiến trúc YOLO: base network (Darknet Architecture) có tác dụng trích suất các đặc trưng. Output của base network là một feature map có kích thước 7x7x1024 sẽ được sử dụng làm input cho các Extra layers. Extra Layers được áp dụng để phát hiện vật thể trên feature map của base network.



(a)



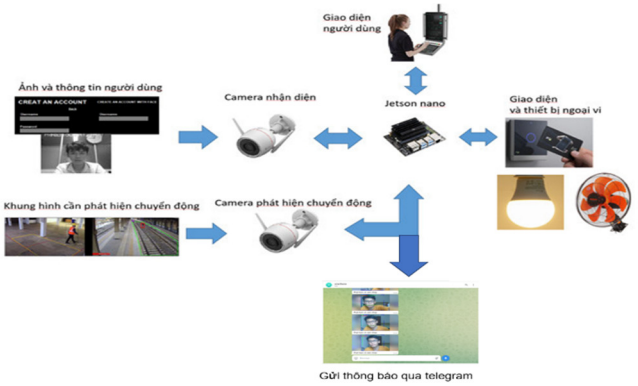
(b)

Hình 9. (a) Sơ đồ kiến trúc mạng YOL, (b) Sơ đồ khối quá trình cảnh báo xâm nhập

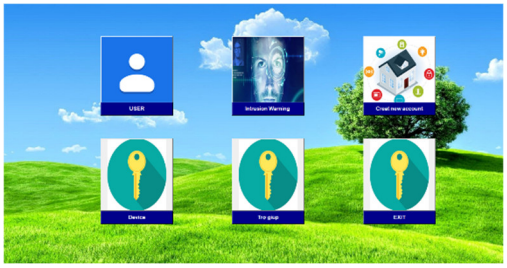
YOLOv3 (You Only Look Once version 3) là một mô hình nhận dạng vật thể tiên tiến và hiệu quả. Ưu điểm của YOLOv3 là việc sử dụng bộ dữ liệu COCO (Common Objects in Context) để huấn luyện mô hình, đây là một tập dữ liệu lớn chứa thông tin về 80 lớp vật thể khác nhau trong nhiều ngữ cảnh. Trong hệ thống smart home Yolov3 được sử dụng để nhận diện đối tượng (cụ thể ở đây là con người) trong phạm vi có bán kính xác định.

3. KẾT QUẢ VÀ THẢO LUẬN

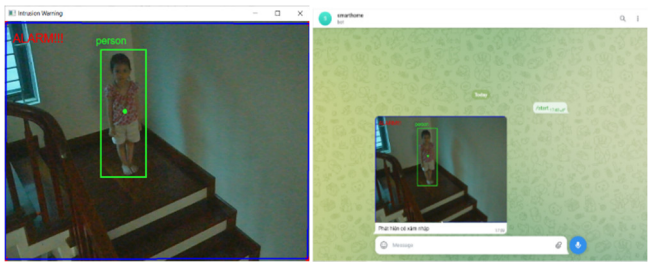
Mô hình hệ thống nhà thông mình IoT sẽ hoạt động trên máy tính nhúng Jetson Nano 8GB. Thông tin của người dùng sẽ được lưu vào cơ sở dữ liệu của hệ thống. Hệ thống gồm an ninh nội vi với dữ liệu đầu vào là ảnh user cho nhận diện, an ninh ngoại vi cho phát hiện đối tượng di chuyển trong khu vực giám sát quanh nhà. Máy tính nhúng Jetson Nano sẽ xử lý và lưu trữ dữ liệu. Người dùng đăng nhập thành công được cấp quyền điều khiển các thiết bị ngoại vi trong nhà.



Hình 10. Cấu trúc hệ thống



(a)

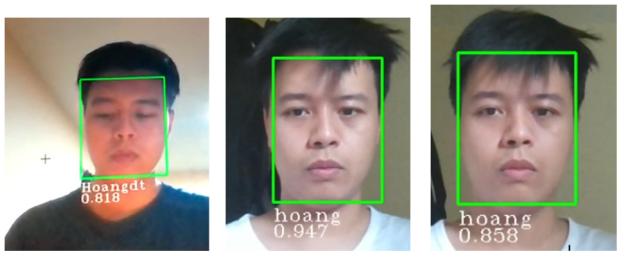


(b) (c)

Hình 11. Giao diện hệ thống (a), phát hiện và thông báo qua telegram (b),(c)

So với kết quả thu được khi sử dụng kỹ thuật Haar Cascades để nhận diện khuôn mặt và chụp ảnh ngay lập tức, bộ ảnh thu được sẽ chỉ bao gồm một đến hai góc khác nhau của khuôn mặt, và sẽ nhận diện nhầm ảnh chụp là người thật. Việc sử dụng thêm phương pháp chụp nhiều góc độ của khuôn mặt

dựa theo tính toán góc nghiêng có thể mang kết quả chính xác hơn, do quá trình training có thể tạo ra nhiều vector có nhiều đặc điểm khuôn mặt hơn, phù hợp trong môi trường bên ngoài (ánh sáng, góc chụp, …) tăng khả năng tìm ra các vector có tính năng tương đương.



(a) (b) (c)

Hình 12. Chất lượng của nhận dạng khuôn mặt với phương pháp lấy mẫu khuôn mặt: (a) chưa có kết hợp Head pose, (b) có kết hợp Head pose với số đối tượng là 3 người (c) có kết hợp Head pose với số đối tượng là 9 người

Phương pháp lấy mẫu mới kết hợp với tính toán góc nghiêng khuôn mặt có tác dụng làm tăng độ chính xác, với số lượng mẫu là 40 thì FPS có thể duy trì 20 -24 FPS. Nhưng nếu tăng số đối tượng cần nhận diện lên, tỉ lệ có xu hướng giảm nhẹ. Kết quả thể hiện ở bảng 1 và 2.

Bảng 1. Bảng so sánh giữa các phương pháp nhận dạng theo độ chính xác và số khung hình/giây (FPS)

Mô hình	Độ chính xác (%)	Số khung hình/giây (FPS)
LBP [4-6]	84-88	19-21
DLIB [4-6]	52-59	9-10
MTCNN+Facenet + Head Pose	90-94	21-25

Bảng 2. Bảng so sánh giữa phương pháp nhận dạng theo độ chính xác, số lượng ảnh mẫu và số đối tượng

Mô hình	Số lượng ảnh mẫu	Số đối tượng	Độ chính xác
MTCNN	10	3	81 %
MTCNN	40	3	94 %
MTCNN+Facenet+Head Pose	40	9	86 %

4. KẾT LUẬN

Kỹ thuật nhận diện khuôn mặt bằng kỹ thuật tính toán góc nghiêng kết hợp với mô hình Facenet làm tăng hiệu quả nhận diện cho thuật toán, tăng lớp bảo mật và giúp ngăn chặn một số phương thức giả danh hiện nay. Độ chính xác đạt 90 - 95% với tốc độ xử lý hình ảnh 20 - 24 FPS, mặc dù có xu hướng giảm nếu tăng lên nhiều mẫu những kết quả khả quan so với các phương pháp nhận diện trực quan trong các công trình công bố trước đây. Kỹ thuật đã được tích hợp vào hệ thống IoT và bảo mật nhằm nâng cao tính bảo mật cho ngôi nhà thông minh đồng thời ứng dụng để điều khiển một số thiết bị trong ngôi nhà. Kết quả phần nhận diện khuôn mặt có thể kết hợp điều khiển, giám sát thiết bị điện, tích hợp lên điện thoại thông minh, giám sát và phân tích các dữ liệu của thời tiết môi trường và đưa ra những tín hiệu khẩn cấp nhằm đối phó với những tình huống bất ngờ xảy ra.

TÀI LIỆU THAM KHẢO

[1]. Dang T. V., Bui N. T., "Multi-scale Fully Convolutional Network based Semantic Segmentation for Mobile Robot Navigation," *Electronics*, 12(3), 533, 2022.

[2]. Dang T. V., "Smart home Management System with Face Recognition based on ArcFace model in Deep Convolutional Neural Network," *Journal of Robotics and Control (JRC)*, 3 (6), 754-761, 2022.

[3]. Dang T. V., Bui N. T., "Design the abnormal object detection system using template matching and subtract background algorithm," *MMMS 2022, LNME*, 2023.

[4]. Dang T. V., Phan V. T, Nguyen H. T., Hoang G. M., Bui N. T., "Design of a Face Recognition Technique based MTCNN and ArcFace," *MMMS 2022, LNME*, 2023.

[5]. Dang, T. V., 2023. Smart Attendance System based on Improved Facial Recognition. *Journal of Robotics and Control (JRC)*, 4(1), 46-53.

[6]. Dang, T. V., Linh H. T., "A Secured, Multilevel Face Recognition based on Head Pose Estimation, MTCNN and FaceNet," *Journal of Robotics and Control (JRC)*, 4, 4, 2023.

[7]. Midhun P. Mathew, Therese Yamuna Mahesh, "Leaf-based disease detection in bell pepper plant using YOLO v5," *Signal Image and Video Processing*, 16(3), 841-847, 2022.

[8]. Nguyen N. Q., Su S. F., Tran Q. V., Nguyen V. T., Jeng J. T, "Real time human tracking using improved CAM-shift," In *2017 Joint 17th World Congress of International Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems (IFSA-SCIS)*, 1-5, 2017.

[9]. Nguyen V. T., Nguyen A. T., Nguyen V. T., Bui H. A., "A real-time human tracking system using convolutional neural network and particle filter," In *Intelligent Systems and Networks: Selected Articles from ICISN 2021*, Vietnam, 411-417, 2021.

AUTHORS INFORMATION

Ta Trong Khanh, Dang Thai Viet, Duong Thai Hoang, Vi Trung Kien

Hanoi University of Science and Technology, Vietnam