ĐIỀU KHIỂN CỬA THÔNG MINH SỬ DỤNG NHẬN DẠNG KHUÔN MẶT

INTELLIGENT DOOR CONTROL USING FACE RECOGNITION

Đoàn Thị Hương Giang[1,*]

TÓM TẮT

Tự động hóa và trí tuệ nhân tạo ngày càng phát triển trên toàn thế giới. Ứng dụng các thành tựu của khoa học công nghệ diễn ra trong mọi lĩnh vực của đời sống hàng ngày nhằm đem lại sự tiện lợi, an toàn, bảo mật và riêng tư. Cửa tự động sử dụng các kỹ thuật của tự động hóa để có thể giúp con người tác động điều khiển các thành phần cơ khí của cửa một cách dễ dàng thông qua các thiết bị như động cơ, khóa điện từ, cảm biến, bo mạch điện tử,… mà không cần phải dùng lực của con người để mở cửa. Tuy nhiên, vấn đề bảo mật của cửa tự động thường thực hiện thông qua hệ thống phím để nhập mã hoặc sử dụng thẻ từ. Đây là một trong những cách thức rất dễ bị đánh cắp thông tin. Ngày nay, cùng với sự phát triển của thị giác máy tính và trí tuệ nhân tạo thì một phương thức mới nhằm giúp thông minh hóa thiết bị điều khiển giúp bảo mật cửa thông qua thông tin hình ảnh của khuôn mặt. Trong bài báo này, tác giả đề xuất một hệ thống điều khiển cửa thông minh sử dụng nhận dạng khuôn mặt hoàn thiện có thời gian đáp ứng nhanh và độ chính xác ổn định. Ngoài ra, bài báo cũng đề xuất cách thức chống giả mạo thông qua một cơ chế tương tác người dùng trên thông tin hình ảnh khuôn mặt cũng như kết hợp với bảo mật theo tầng để tăng tính an toàn khi sử dụng thông tin hình ảnh.

Từ khóa: *Cửa tự động, học sâu, học máy, phát hiện khuôn mặt, nhận dạng khuôn mặt, tương tác người - máy, cửa thông minh.*

ABSTRACT

Automation and Artificial Intelligence have grown over the world. The achievements of science and technology has applied in all fields of our daily life in order to bring convenience, safety, security and privacy. Automatic door uses the techniques of automation to help end-users automatically control the mechanical components of the door through devices such as motors, electromagnetic locks, sensors, and electronic circuits,... without using human force to open doors. However, the traditional automatic door is securitized through the key system to enter codes or use special magnetic cards. This is one of the most vulnerable ways to get information stolen. Nowadays, the development of Computer Vision and Artificial Intelligence that brings a new advantage method to control devices, secure the door through image information of human face. In this paper, we propose an intelligent door that could be controlled by facial recognition system with real-time, robust system and stable accuracy. In addition, this paper also proposes how to prevent tampering through a user interaction mechanism based on facial image information as well as combining with security cascade to increase safety using face cues.

Keywords**:** *Automatic door, deep learning, machine learning, face detection, face recognition, human-machine interaction, intelligent door.*

[1]Khoa Điều khiển và Tự động hóa, Trường Đại học Điện lực

[*]Email: giangdth@epu.edu.vn

Ngày nhận bài: 30/8/2021

Ngày nhận bài sửa sau phản biện: 30/9/2021

Ngày chấp nhận đăng: 25/10/2021

1. GIỚI THIỆU

Ngày nay, bài toán nhận dạng khuôn mặt tập trung sự quan tâm của nhiều nhà khoa học [6-11] bởi đây là một trong những thông tin đặc thù và riêng biệt của con người và có thể khai thác vào nhiều ứng dụng khác nhau như tương tác người máy (Human Machine-Interaction: HCI) [6], định danh [7], định danh lại [8], điều khiển và tự động hóa [9], bảo mật [10], giám sát [11]…. Tuy nhiên, bài toán này cũng còn đang phải đối mặt với một số thách thức khi triển khai hệ thống thực, tính bảo mật hệ thống, độ chính xác tùy ứng dụng,… Bài toán nhận dạng khuôn mặt có thể chia thành các bước như: Phát hiện khuôn mặt, biểu diễn khuôn và nhận dạng. Ở bước đầu tiên là phát hiện khuôn mặt đã có nhiều phương pháp được đề xuất như Haarlike Cascade [12], Dlib [1] hay phương pháp sử dụng các mạng học sâu YoloV3 [4], MTCNN [17, 18]. Trong bài báo [6], các tác giả đã có thử nghiệm định tính trên ba phương pháp trên. Tuy nhiên, các tác giả mới chỉ thử nghiệm trên bộ cơ sở dữ liệu (CSDL) tự thu thập mà chưa đánh giá trên một bộ CSDL của cộng đồng dùng chung. Do đó, trong nghiên cứu này nhóm tác giả sử dụng thêm YoloV4 [5] và không những chỉ đánh giá các phương pháp này trên bộ CSDL tự thu thập mà còn đánh giá trên bộ CSDL khác nữa để có cái nhìn tổng quan hơn. Trong khâu biểu diễn khuôn mặt, tác giả sẽ sử dụng các giải pháp biểu diễn điểm đặc trưng có đề xuất kết hợp giải pháp chống giả mạo dựa trên một số khung hình liên tiếp nhau. Khâu nhận dạng khuôn mặt tác giả sẽ sử dụng bộ phân lớp SVM [3] để nhận dạng và đánh giá sự hiệu quả của phương pháp biểu diễn. Tuy nhiên, với phương pháp sử dụng bộ phân lớp trong điều khiển học sẽ gặp phải một số vấn đề về bảo mật, do với mỗi khuôn mặt đưa vào mô hình đã huấn luyện, bộ phân lớp luôn phân chúng vào một trong số các mô hình lớp đã được huấn luyện trước đó. Hoặc nếu để không bị nhận nhầm thì phải đảm bảo luôn có một lớp chứa rất nhiều mẫu không đúng. Điều đó là rất khó khăn đối với các bài toán thực tế. Do đó, trong nghiên cứu này, tác giả sau khi đã đánh giá hiệu quả của đặc trưng biểu diễn sẽ xem xét đánh giá lại về sự sai khác của mẫu với mô hình kết hợp với thang đo RMSE[15]. Mô hình hệ thống sau đó sẽ được triển khai và đánh giá hệ thống điều khiển thực tế cả về độ chính xác và thời gian đáp ứng hệ thống.

Phần tiếp theo của bài báo gồm các phần sau đây: Phần 2 mô tả chi tiết giải pháp đề xuất. Kết quả thử nghiệm được

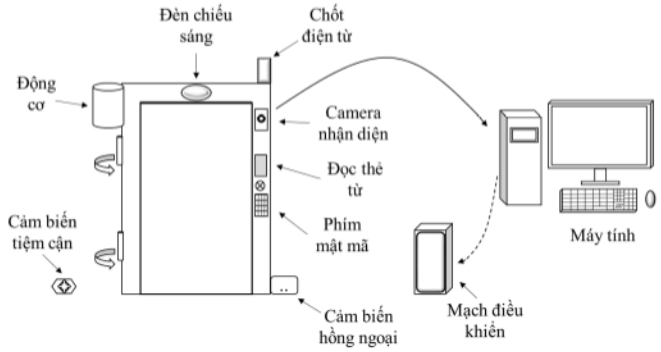
trình bày trong phần 3. Phần 4 là mục cuối cùng sẽ trình bày kết luận và hướng phát triển trong thời gian tiếp theo

2. GIẢI PHÁP ĐỀ XUẤT

Hệ thống cửa thông minh đề xuất gồm ba phần chính: (1) Thiết kế phần cứng bộ điều khiển cửa. Trong đó, phần cứng sử dụng vi điều khiển Adruino là bo mạch trung tâm chi tiết được trình bày trong mục 2.1; (2) Phần mềm điều khiển cửa; (3) Phần mềm nhận dạng thông tin hình ảnh khuôn mặt qua camera RGB thường và sử dụng máy tính PC cá nhân được trình bày trong mục 2.3. Trong đó, phần mềm nhận dạng khuôn mặt được mô tả như trong hình 1 với hai bước chính là: Thu thập cơ sở dữ liệu và huấn luyện hệ thống. Sau đó, khâu phát hiện và nhận dạng khuôn mặt có tích hợp chức năng chống giả mạo được thực hiện trực tuyến. Các nội dung được trình bày chi tiết trong các mục tiếp theo của bài báo.

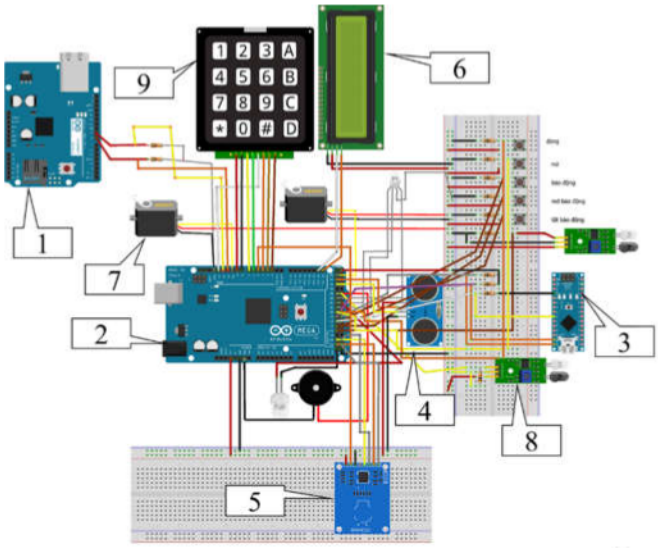
Hình 1. Sơ đồ khối hệ thống nhận dạng khuôn mặt

2.1. Ghép nối phần cứng



Hình 2. Cấu trúc ghép nối hệ thống

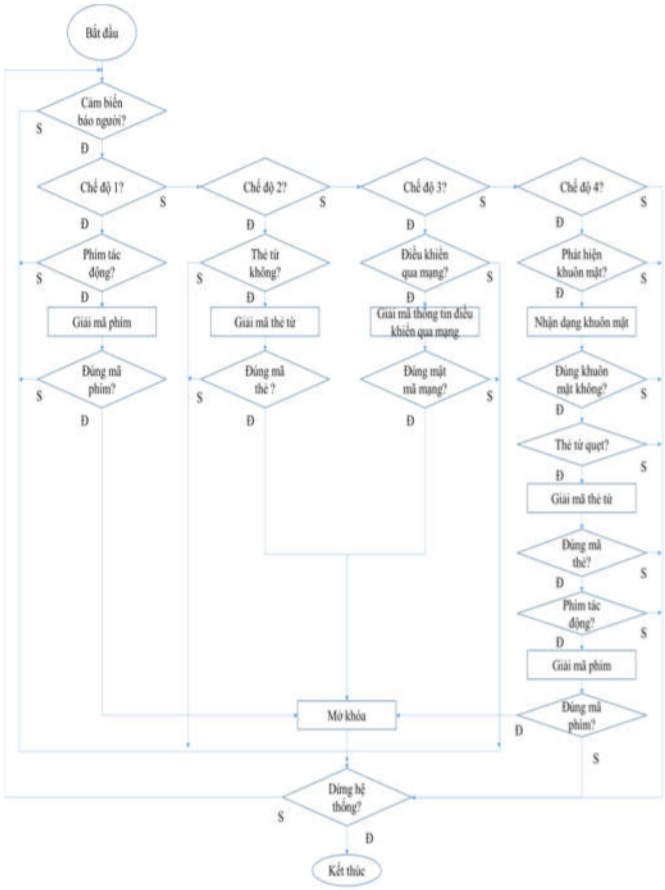
Cấu trúc phần cứng của hệ thống cửa được mô tả như trong hình 2. Trong đó, cửa là loại mở cánh, có một khóa chốt điện từ loại LY- 01 giúp khóa cửa, cảm biến tiệm cận tác động theo mức để xác định trạng thái hiện tại của cửa đã đóng hay đang mở cửa, động cơ mở cửa sử dụng là loại động cơ servo cho phép điều chỉnh chính xác vị trí góc mong muốn dùng để mở hoặc đóng cửa sau khi chốt đã được mở khóa. Cảm biến hồng ngoại để xác định có người ở cửa hay không. Mạch đọc thẻ từ và hệ thống phím là hai phương thức có thể hoạt động độc lập hoặc tham gia vào làm một khâu trong chế độ đa thể thức. Đèn chiếu sáng nhằm hỗ trợ cho hệ thống camera có được hình ảnh rõ nét nhất. Mạch điều khiển nhóm tác giả sử dụng là Arduino mega 2560 [13] để nhận lệnh điều khiển từ Internet và máy tính thông qua kết nối với mạch Arduiro Ethernet Shield W5100 [14] - mạch giúp kết nối mạng với máy tính. Trong đó, mạch Arduino mega 2560 sẽ có tác dụng điều khiển ra các cơ cấu chấp hành của động cơ chốt cửa và động cơ mở cửa. Sơ đồ đi dây và đấu nối phần cứng của mạch điều khiển được minh họa như trong hình 3.



Hình 3. Sơ đồ ghép nối phần cứng mạch điều khiển

1. Arduino Uno R3; 2. Arduino Mega2560; 3. Arduino Nano; 4. Cảm biến HC-SR04; 5. Module RFID; 6. LCD và I2C; 7. Động cơ servo; 8. Cảm biến hồng ngoại MH-IR01; 9. Ma trận phím 4x4

2.2. Xây dựng hệ thống điều khiển ở chế độ bảo mật cao



Hình 4. Lưu đồ thuật toán các chế độ điều khiển của cửa thông minh

Hệ thống cửa được tác giả thiết kế với bốn chế độ. Trong đó có ba chế độ điều khiển đơn phương thức (chế độ

1,2,3) gồm: mật mã thông qua hệ thống khóa số, thẻ từ và webserver. Đây là những giải pháp thực hiện nhằm nội địa hóa sản phẩm cửa thông minh. Tuy nhiên, trong nghiên cứu này tác giả sẽ nghiên cứu và đề xuất giải pháp kết hợp đa thể thức (chế độ 4) nhằm kết hợp cả ba loại phương thức đều khiển ở trên với thông tin nhận dạng khuôn mặt. Đây là giải pháp có khả năng bảo mật cao hơn hẳn các chế độ sử dụng các phương thức khác do mỗi người sẽ có một đặc điểm khuôn mặt hoàn toàn khác nhau. Tuy nhiên, ngay cả với thông tin khuôn mặt thì vẫn có thể bị giả mạo khi kẻ gian chỉ cần một tấm ảnh của gia chủ. Do đó, tác giả đề xuất giải pháp nhận dạng khuôn mặt có sử dụng chống giả mạo trên thông tin hình ảnh. Giải pháp kết hợp đa thể thức được xem là chế độ mật cao sử dụng giải pháp nối tầng cả bốn loại phương thức. Quá trình hoạt động của các chế độ trên được thể hiện như lưu đồ thuật toán ở hình 4.

Chế độ 4 là giải pháp điều khiển cửa đảm bảo tính bảo mật cao nhất trong bốn chế độ. Trong đó, chế độ này được bảo mật thông qua nối tầng liên tiếp nhiều phương thức. Cửa được mở qua ba bước liên tiếp gồm: Kiểm tra thông tin khuôn mặt có chống giả mạo; Kiểm tra thông tin thẻ; Kiểm tra mã khóa số và Điều khiển qua mạng LAN. Bước đầu tiên của hệ thống là thu nhận hình ảnh khuôn mặt của đối tượng bằng camera giám sát được lắp đặt tại cửa mỗi căn hộ. Hệ thống tiến hành phân tích tại khối xử lý trung tâm vừa để nhận dạng và đồng thời có kiểm tra giả mạo thông tin khuôn mặt. Sau đó, tín hiệu đúng sẽ được truyền tín hiệu xuống Arduino thông qua mạng LAN. Bước tiếp theo, hệ thống sẽ yêu cầu quẹt thẻ từ thông qua module RFID RC522 và nhập mật khẩu ở bàn phím thông qua chuỗi mã hóa đã được thiết lập trước đó bởi người sử dụng. Với trường hợp nhập mật khẩu thành công, hệ thống điều khiển Arduino đưa tín hiệu mở rơ le cấp nguồn cho khóa điện từ thực hiện mở chốt cửa. Sau thời gian đặt trước, để đảm bảo chốt điện từ đã được mở thì động cơ mở cửa sẽ hoạt động. Ngược lại, khi nhập mật khẩu hoặc giải mã thẻ từ thất bại nhiều hơn số lần hệ thống yêu cầu thì hệ thống sẽ gửi tin nhắn và gọi điện cảnh báo về số điện thoại đã được lưu trước đó trên hệ thống. Trong trường hợp nhận diện khuôn mặt không thành công tại bước thứ nhất, hệ thống hủy bỏ lệnh nhập mật khẩu hay giải mã thẻ từ RFID RC 522. Trong nghiên cứu này, hệ thống điều khiển cửa được trang bị thêm nút bấm ưu tiên giúp cho việc mở cửa thuận tiện hơn trong những trường hợp đặc biệt khi chủ nhà không yêu cầu tiến hành kiểm tra bảo mật.

Việc đánh giá hoạt động của quá trình bảo mật này sẽ được tác giả thực hiện thử nghiệm trong mục 2.3 sau đây. Ở bước đầu tiên của chế độ bảo mật cao được thực hiện sử dụng kết quả của quá trình nhận dạng khuôn mặt. Chi tiết của hệ thống nhận dạng khuôn mặt được trình bày trong mục 2.3 sau đây.

2.3. Nhận dạng khuôn mặt

2.3.1. Phát hiện khuôn mặt

Do ảnh thu thập thường không chỉ chứa mỗi khuôn mặt và có thể có thêm thông tin về cả người hoặc nền. Do đó, để loại bớt thông tin dư thừa gây nhiễu và làm giảm độ chính xác cũng như làm tăng thời gian của quá trình nhận dạng khuôn mặt. Trong khuôn khổ của nghiên cứu này, tác giả sử dụng phương pháp phát hiện khuôn mặt bằng Dlib. Đây là phương pháp được đề xuất bởi các tác giả trong bài báo [1] và cung cấp mã nguồn cũng như mô hình đã huấn luyện dưới định dạng file.xml cho cộng đồng nghiên cứu tại [2]. Dlib là một phương pháp giúp phát hiện khuôn mặt chính xác hơn Haar like Cascade và thời gian đáp ứng khá nhanh. Chi tiết đánh giá độ chính xác và thời gian đáp ứng của mô hình khi triển khai hệ thống được thể hiện chi tiết trong mục 3. Với mỗi khung hình khi thu thập từ camera (I_{RGB}), sau khi qua Dlib sẽ lấy được các thông tin của vùng chữ nhật bao quanh khuôn mặt như công thức (1) sau đây:

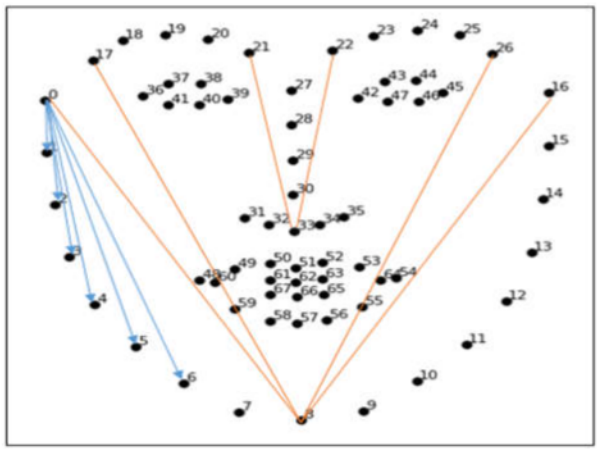
$$REC_{face}(x, y, w, h) = F_{Dlib}(I_{RGB}) \quad (1)$$

Trong đó, x và y là tọa độ theo trục x (trục ngang ảnh) và y (trục dọc ảnh) của đỉnh trên cùng bên trái của vùng khuôn mặt phát hiện được, w là độ rộng của vùng khuôn mặt, h là chiều cao của vùng khuôn mặt. Sau đó, vùng khuôn mặt được cắt dựa trên các tham số trên của ảnh màu (RGB) như công thức (2) sau đây:

$$I_{face} = F_{Cropt}(I_{RGB}, REC_{face}) \quad (2)$$

2.3.2. Biểu diễn khuôn mặt

Hình ảnh khuôn mặt (I_{face}) sau khi đã được phát hiện và cắt loại bỏ thông tin dư thừa như trình bày trong mục trước sẽ được trích chọn đặc trưng để nhận dạng. Mặc dù hiện nay có nhiều giải pháp trích chọn đặc trưng đã được đề xuất như [1, 6]. Tuy nhiên, trong khuôn khổ của nghiên cứu này, với Dlib [1, 2], ngoài việc có thể phát hiện chính xác vùng khuôn mặt trên ảnh thì phương pháp này còn giúp phát hiện ra 68 điểm đặc trưng trên khuôn mặt (Pi(i = 0 ÷ 67)) với thời gian đáp ứng nhanh và chính xác. Trong đó, các điểm bao quanh hàm là từ P_0 đến P_{17}; lông mày trái từ điểm P_{17} đến P_{21}; lông mày phải từ điểm P_{22} đến P_{26}; mũi từ điểm P_{27} đến P_{35}; mắt trái từ điểm P_{36} đến P_{41}; mắt phải từ điểm P_{42} đến P_{45} và miệng từ điểm P_{48} đến P_{67}. Trong khuôn khổ nghiên cứu này, tác giả đề xuất giải pháp biểu diễn khuôn mặt sử dụng các điểm đặc trưng như biểu diễn trong hình 5.



Hình 5. Biểu diễn đề xuất các đoạn thẳng và góc trên khuôn mặt

Mỗi đoạn thẳng màu xanh trên hình vẽ được định nghĩa như sau:

$$d_{i+1} = \sqrt{(x_{i+1} - x_0)^2 + (y_{i+1} - y_0)^2} (i = [0 \div 66]) \quad (3)$$

Trong đó: x_0, y_0 là tọa độ của điểm 0. x_{i+1}, y_{i+1} là tọa độ của các điểm lần lượt từ điểm 1 đến điểm 67. Véc tơ đặc trưng biểu diễn cho khuôn mặt sử dụng các đoạn thẳng nối bởi các điểm tiêu biểu trên mỗi khuôn mặt được biểu diễn như công thức (4):

$$F_{68points}^{distance} = [d_1; d_2; \dots; d_{67}] \quad (4)$$

Ngoài ra, trên khuôn mặt có các thông tin của một số điểm là cố định và khác biệt giữa người này với người kia. Tác giả nhận thấy các góc như biểu diễn bởi các cặp đoạn thẳng màu cam trong hình 5 ít thay đổi trong ngữ cảnh bài toán điều khiển cửa mà tác giả đề xuất. Do đó, các góc được tính toán và sử dụng để biểu diễn thông tin của khuôn mặt như trong công thức (5):

$$\cos(\theta_l) = \frac{\overrightarrow{n_1^l} * \overrightarrow{n_2^l}}{|n_1^l| * |n_2^l|} \quad (5)$$

Véc tơ đặc trưng cho góc được biểu diễn bởi công thức (6):

$$F_{angle} = [\cos\theta_0, \cos\theta_1, \dots, \cos\theta_m] \quad (6)$$

Để các tọa độ là bất biến với các vị trí xa hay gần của đối tượng và với camera, trong bài báo này chúng tôi sử dụng tỉ lệ của các đoạn thẳng so với tọa độ điểm đầu tiên như biểu diễn trong công thức (7):

$$r_i = \frac{d_i}{d_1} \quad (7)$$

Trong đó: d_1 là khoảng cách từ điểm 0 đến điểm 1; d_i là khoảng cách từ điểm 0 đến điểm i, $i = [1 \div 67]$. Véc tơ đặc trưng biểu diễn cho tỉ lệ đoạn thẳng được biểu diễn như công thức (8):

$$F = [r_1; r_2; \dots; r_{67}] \quad (8)$$

Véc tơ đặc trưng tổng hợp biểu diễn cho khuôn mặt sử dụng kết hợp đặc trưng cạnh và đặc trưng góc được biểu diễn bởi công thức (9):

$$F = \left[\|F_{angle}\|_2; \|F\|_2\right] \quad (9)$$

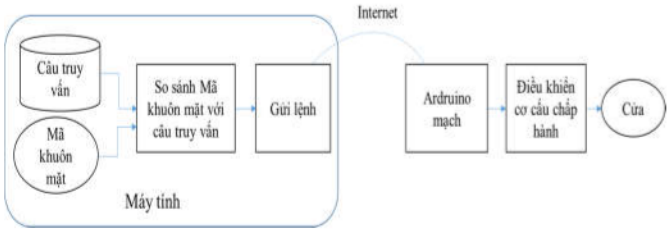
2.3.3. Nhận dạng

- Sử dụng bộ phân lớp SVM: Trong bài báo này, các véc tơ đặc trưng trong các công thức (4), (6) và (7) sẽ lần được coi là đầu vào cho bộ phân lớp SVM [3] để nhận dạng. Đầu ra của bộ phân lớp sẽ là mã của người được được gán nhãn trước đó. Với việc sử dụng bộ phân lớp SVM đạt độ chính xác tốt và được thể hiện trong rất nhiều nghiên cứu về phân loại đối tượng. Tuy nhiên, phương pháp này gặp phải vấn đề về phân lớp nhầm vào lớp có mẫu gần đúng nhất, tức là khuôn mặt gần đúng nhất trong tập dữ liệu huấn luyện. Do đó, trong khuôn khổ bài báo này, tác giả sẽ kết hợp đầu ra của SVM với độ đo khoảng cách RMSE để loại trừ các lỗi nhận nhầm gần đúng của phương pháp SVM.

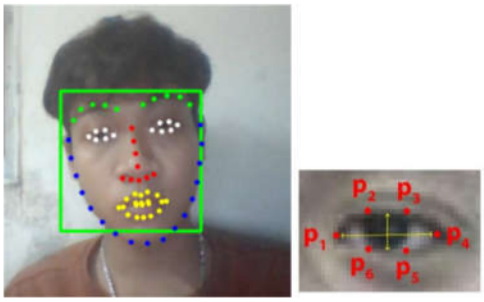
- Sử dụng độ đo khoảng cách RMSE: Các véc tơ đặc trưng của khuôn mặt như đã trình bày trong mục 2.3.1 và 2.3.2 sẽ được sử dụng để so khớp hay để đánh giá sự sai khác nhau giữa ảnh mới được đưa vào và ảnh sử dụng làm tập mẫu. Tác giả sử dụng thang đo RMSE để tính toán sự sai khác giữa các véc tơ đặc trưng của dữ liệu với véc tơ đặc trưng của tập mẫu. Cụ thể, nếu có véc tơ đặc trưng của tập mẫu là $\mathbf{T} = [\mathbf{P_i^1}(\mathbf{x}, \mathbf{y}) \quad \dots \quad \mathbf{P_i^n}(\mathbf{x}, \mathbf{y})]$ và véc tơ đặc trưng của tập thử nghiệm là $\mathbf{T^*} = [\mathbf{P_j^1}(\mathbf{x}, \mathbf{y}) \quad \dots \quad \mathbf{P_j^n}(\mathbf{x}, \mathbf{y})]$. Khoảng cách RMSE được tính theo công thức (10):

$$\mathrm{RMSE}(T, T^*) = \sqrt{\frac{\sum_{k=1}^{n}\left(P_i^k(x,y) - P_j^k(x,y)\right)^2}{n}} \quad (10)$$

2.3.4. Chống giả mạo khuôn mặt



Hình 6. Quá trình chống giả mạo sử dụng thông tin hình ảnh



Hình 7. Các điểm đặc trưng trên khuôn mặt và 6 điểm đặc trưng trên mắt

Hình 6 thể hiện quá trình chống giả mạo với các câu truy vấn được gia chủ cài đặt. Nếu với mỗi câu truy vấn mà biểu cảm khuôn mặt khớp nhau thì việc so khớp được trả về kết quả đúng và thực hiện gửi lệnh điều khiển từ máy tính xuống thiết bị điều khiển. Quá trình kiểm tra mã truy vấn được thực hiện thông qua khả năng chớp mắt của người. Trong đó, mỗi mắt được biểu thị bằng sáu tọa độ p_1, p_2, p_3, p_4, p_5, p_6 được minh họa trong hình 7. Công thức phản ánh mối quan hệ giữa các điểm trên mắt được gọi là tỷ lệ khung hình của mắt (EAR) như biểu diễn trong công thức (11):

$$\mathbf{EAR} = \frac{||\mathbf{p_2} - \mathbf{p_6}|| + ||\mathbf{p_3} - \mathbf{p_5}||}{2 * ||\mathbf{p_4} - \mathbf{p_1}||} \quad (11)$$

Trong đó, $\mathbf{p_1}, \mathbf{p_2}, \mathbf{p_3}, \mathbf{p_4}, \mathbf{p_5}, \mathbf{p_6}$ là các vị trí mốc mắt trên khuôn mặt 2D. Tử số của phương trình này được tính bằng khoảng cách giữa các mốc mắt dọc ($\mathbf{p_2}, \mathbf{p_3}, \mathbf{p_5}, \mathbf{p_6}$), trong khi đó mẫu số được tính bằng khoảng cách giữa các mốc mắt ngang ($\mathbf{p_1}, \mathbf{p_4}$). Tác giả nhận thấy, cần nhân đôi mẫu số bởi vì chỉ có một tập hợp các điểm nằm ngang nhưng có hai tập hợp các điểm thẳng đứng. Tỷ lệ khung hình của mắt gần như không đổi khi mắt đang mở, nhưng sẽ nhanh chóng giảm xuống 0 khi diễn ra quá trình chớp mắt. Từ đó, ta dựa vào tỷ lệ khoảng cách **EAR** để xác định xem một người có đang chớp mắt hay không. Qua đó, kiểm chứng được đối tượng có

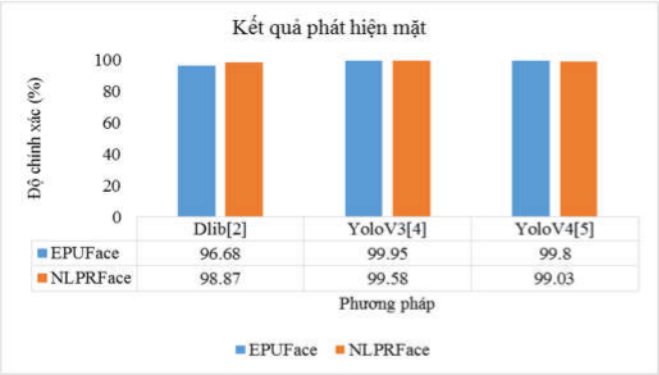
phải là người thật hay không. Bằng cách khi phát hiện đối tượng sẽ bắt đầu kiểm tra nháy máy. Nếu nháy mắt ba lần trong thời gian quy định thì đó là người thật. Ngược lại, kết luận đó là ảnh tĩnh được đưa vào nhận dạng.

3. KẾT QUẢ THỬ NGHIỆM

Trong mục này, tác giả sẽ trình bày một số kết quả thử nghiệm của giải pháp đề xuất. Các thử nghiệm được thực hiện trên máy tính PC Core i5 3.l GHz CPU, 4GB. Ngôn ngữ lập trình để thực hiện quá trình thu thập, nhận dạng khuôn mặt và kết nối mạng là Python. Ngôn ngữ lập trình để viết chương trình cho Adruino là C++. Các thử nghiệm trên khuôn mặt được thực hiện gồm hai pha gồm: pha trực tuyến và pha không trực tuyến. Ở pha không trực tuyến tác giả sử dụng CSDL tự thu thập EPUFace[6] và cơ sở dữ liệu của cộng đồng dùng chung NLPRFace[16]. Ở pha trực tuyến tác giả sử dụng hình ảnh khuôn mặt sinh viên của trường Đại học Điện lực và thử nghiệm bằng camera của máy tính xách tay.

Các thử nghiệm được tiến hành trong nghiên cứu này gồm: (1) Đánh giá quá trình phát hiện khuôn mặt trên các giải pháp khác nhau; (2) Đánh giá kết quả nhận dạng khuôn mặt; (3) Đánh giá độ chính xác và thời gian đáp ứng toàn hệ thống; (4) Đánh giá hiệu quả của hệ thống chống giả mạo khuôn mặt.

3.1. Kết quả phát hiện khuôn mặt



Hình 8. Kết quả phát hiện khuôn mặt

Hình 8 là kết quả độ chính xác của quá trình phát hiện khuôn mặt và bảng 1 thể hiện thời gian tương ứng để nhận dạng. Đánh giá được thực hiện trên hai bộ CSDL. Ba phương pháp được triển khai nhận dạng gồm Dlib[2], YoloV3 [4] và YoloV4 [5]. Hai tham số được sử dụng gồm độ chính xác phát hiện và thời gian đáp ứng của quá trình phát hiện. Thử nghiệm được thực hiện trên hai bộ CSDL gồm: (1) bộ CSDL EPUFace [6, 19] gồm 10 người, mỗi người 200 ảnh với độ phân giải 640x480 điểm ảnh; (2) Bộ CSDL NLPRFace [16] gồm 450 ảnh của 27 đối tượng với độ phân giải 896x592 điểm ảnh. Kết quả trên cho thấy rằng, phương pháp Dlib có kết quả nhận dạng thấp hơn (đạt 96,68% và 98,87% trên hai bộ CSDL), so với hai phương pháp sử dụng mạng học sâu còn lại thì phương pháp YoloV3 [4] đạt kết quả cao nhất lên tới 99,95% và 99,58%. Tuy nhiên, thời gian đáp ứng của phương pháp Dlib lại thấp nhất và thấp hơn rất nhiều so với hai phiên bản của Yolo với chỉ 66ms cho bộ CSDL EPUFace và 71ms cho bộ CSDL NLPRFace. Để triển khai ứng dụng thực thế thì yếu tố thời gian là một trong những thông số đáng qua tâm. Do đó, trong nghiên cứu này, tác giả sẽ sử dụng phương pháp Dlib để triển khai ứng dụng thực tế.

Bảng 1. Thời gian phát hiện khuôn mặt

	Dlib [2]	YoloV3 [4]	YoloV4 [5]
EPUFace	0,066ms	0,7ms	2,9ms
NLPRFace	0,071ms	0,67ms	3,1ms

3.2. Kết quả nhận dạng khuôn mặt

Trong bảng 2, tác giả sử dụng kết quả phát hiện khuôn mặt sử dụng phương pháp Dlib [2]. Kết quả nhận dạng khuôn mặt sử dụng bộ phân lớp SVM [3] trên bộ CSDL EPUFace [6, 19] và NLPRFace [16] được thể hiện trong bảng 2, tác giả đã tiến hành thử nghiệm với bốn hàm nhân và đưa ra kết quả với từng hàm nhân tương ứng. Trong đó, nhận dạng khuôn mặt đạt kết quả cao nhất là 97,14% (tương đương với phương pháp tác giả đã đề xuất trong [6]) và 99,72% cho hai bộ CSDL tương ứng. Ngoài ra, kết của còn cho thấy hàm nhân là RBF đạt kết quả cao hơn so với ba hàm nhân còn lại Linear, Sigmoid và Poly với kết quả chỉ đạt 92,13%, 15,75%, 90,94% cho bộ CSDL EPUFace và 96,67%, 13,45% và 86,56% đối với bộ CSDL NLPRFace. Kết quả này thêm một lần nữa khẳng định phương pháp biểu diễn khuôn mặt đề xuất hiệu quả đối bộ phân lớp RBF SVM.

Bảng 2. Kết quả nhận dạng khuôn mặt sử dụng SVM

Hàm nhân SVM	EPUFace		NLPRFace	
	Độ chính xác (%)	Thời gian đáp ứng (ms)	Độ chính xác (%)	Thời gian đáp ứng (ms)
Linear	92,13	71	96,67	69
Sigmoid	15,75	66	13,45	64
Rbf	97,14	71	99,72	70
Poly	90,94	67	86,56	66

Trong bảng 3, tác giả thực hiện đánh giá khi sử dụng thang đo RMSE để nhận dạng khuôn mặt kết hợp với phương pháp SVM. Trong đó, chỉ có những kết quả đã phân lớp đúng trên SVM mới được đưa vào để kiểm tra trên thang đo RMSE để loại trừ các trường hợp nhận nhầm. Kết quả cho thấy độ chính xác của hai bộ CSDL đạt kết quả thấp hơn không đáng kể khi so sánh với kết quả sử dụng bộ phân lớp RBF SVM như ở bảng 2 (phương pháp tốt nhất được khảo sát đối với bộ phân lớp SVM). Thời gian đáp ứng lâu hơn rất nhiều so với sử dụng SVM do quá trình so khớp mẫu được thực hiện trên toàn bộ tập mẫu. Tuy nhiên, khi tác giả thực hiện so khớp ở khâu trực tuyến sẽ được cải tiến như trình bày ở phần cuối mục 3.2. Trong đó, thời gian nhận dạng của hàm nhân RBF đạt được ngang bằng với hàm nhân Linear với 71ms đối với CSDL EPUFace và NLPRFace lần lượt là 69ms và 70ms. Tuy nhiên, khi sử dụng thang đo RMSE, tác giả khảo sát và lấy độ đo khoảng cách để đảm bảo sự tương đồng là 0,08 với cả hai tập dữ liệu. Kết quả phân loại đúng đạt 96,75% trên bộ CSDL EPUFace và lên tới 98,34% trên bộ CSDL NLPRFace. Mặc dù CSDL NLPRFace có nhiều lớp hơn nhưng kết quả nhận dạng

lại cao hơn bộ CSDL EPUFace do độ phức tạp của các mẫu trong bộ CSDL. Trong đó, bộ CSDL NLPRFace chỉ thu thập các khuôn mặt đơn giản, góc chính diện trong khi bộ CSDL EPUFace thu thập tại nhiều vị trí và góc nhìn khác nhau. Với phương pháp kết hợp như đề xuất thì các mẫu được coi là không nhận đúng và có chỉ số RMSE cao hơn 0,08 sẽ được phân vào tổ hợp các mẫu sai. Trong khi đó, với kết quả đúng của bộ phân lớp SVM sẽ được kiểm tra lại trên thang đo RMSE để kiểm tra trường hợp nhận nhầm. Điều này dẫn đến khi điều khiển hệ thống sẽ giảm hiện tượng điều khiển sai. Do đó, trong khuôn khổ ứng dụng này tác giả sẽ sử dụng thang đo RMSE kết hợp với VSM để nhận mẫu đối tượng.

Bảng 3. Kết quả nhận dạng khuôn mặt sử dụng thang đo RMSE

CSDL	Độ chính xác (%)	Thời gian đáp ứng (ms)
EPUFace	96,76	240
NLPRFace	98,34	258

Tuy nhiên, một lưu ý ở bước kiểm tra RMSE trong hệ thống trực tuyến, tác giả chỉ lấy mỗi lớp 09 ảnh mẫu ở ba góc nhìn -45^0, 0^0 và 45^0 với mỗi góc nhìn trong tập huấn luyện lấy ba ảnh tại ba vị trí khoảng cách khác nhau để kiểm tra độ sai khác RMSE nhằm giảm thiểu thời gian nhận dạng của toàn hệ thống. Do vậy, tốc độ của khâu nhận dạng trực tuyến giảm chỉ còn khoảng 5fps. Đây có thể xem là tốc độ chấp nhận được khi triển khai ứng dụng thực tế. Trong khi đó, hệ thống được thử nghiệm với 05 người và mỗi người thực hiện 10 lần thì độ chính xác đạt 97%.

3.3. Kết quả quá trình chống giả mạo

Quá trình chống giả mạo là pha đầu tiên của hệ thống điều khiển cửa thông minh. Trong thử nghiệm, nhóm tác giả thực hiện khảo sát trên 5 người gồm 3 nam và 2 nữ. Mỗi người được hướng dẫn thực hiện trong 10 lần. Trong đó, hệ thống sẽ đánh giá trên hai tiêu chí là tính số lần thực hiện chính xác và thời gian đáp ứng tương ứng với mỗi lần thực thi. Phương pháp triển khai đã đạt được độ chính xác trung bình của cả 5 người đạt 95,4% và thời gian trung bình cho mỗi lần thực hiện của người dùng là 15(s). Với độ chính xác cao và thời gian phát hiện có thể xem là khá nhanh. Phương pháp chống giả mạo đề xuất giúp tăng tính an toàn và bảo mật cho hệ thống ở tầng đầu tiên của hệ thống mở cửa sử dụng thông tin hình ảnh.

4. KẾT LUẬN

Trong bài báo này, tác giả đã thiết kế một hệ thống điều khiển cửa thông minh trong đó kết hợp giữa cửa tự động với công nghệ xử lý ảnh, công nghệ học sâu để phát hiện khuôn mặt. Giải pháp của hệ thống đề xuất còn có khả năng chống giả mạo khuôn mặt. Ngoài ra, hệ thống cũng được cài đặt và thử nghiệm hệ thống để kiểm tra độ chính xác cũng như đánh giá thời gian đáp ứng trong điều khiển hệ thống ở mức tương tác đóng mở mạch điện tử.

Tuy nhiên, hệ thống chưa đánh giá với hệ thống cửa thực tế. Trong thời gian tới, tác giả sẽ kết hợp và so sánh với kỹ thuật học sâu tiên tiến hiện nay (deep learning) trong khâu nhận dạng để nâng cao hơn nữa độ chính xác của quá trình nhận dạng khuôn mặt. Ngoài ra, tác giả sẽ đánh giá thử nghiệm trên hệ thống thực tế và tăng số lượng thử nghiệm người dùng.

LỜI CẢM ƠN

Nghiên cứu này được tài trợ bởi Đề tài NCKH cấp trường Đại học Điện lực: "Điều khiển thiết bị điện tử gia dụng thông minh sử dụng kết hợp công nghệ xử lý ảnh và trí tuệ nhân tạo".

TÀI LIỆU THAM KHẢO

[1]. D. E. King, 2009. *Dlib-ml: A machine learning toolkit*. J. Mach. Learn. Res, pp. 1755–1758.

[2]. http://dlib.net/, 2020.

[3]. C. Burges, 1997. *A Tutorial on Support Vector Machines for Pattern Recognition*. Journal of Data Mining and Knowledge Discovery, Vol. 43, pp. 1-43.

[4]. J. Redmon, A. Farhadi, 2018. *Yolov3: An incremental improvement*. CoRR journal, Vol. abs/1804.02767, pp. 1-6.

[5]. A. Bochkovskiy, C. Y. Wang, H. Y. M. Liao, 2020. *Yolov4: Optimal speed and accuracy of object detection*. ArXiv journal, Vol. abs/2004.10934, pp. 1-17.

[6]. Huong-Giang Doan, Ngoc-Trung Nguyen, 2020. *Realtime Face Recognition for Intelligent Door System*. International Journal of Emerging Trends in Engineering Research, Vol. 8, No. 9, pp. 5226-5232.

[7]. N. R. Borkar, S. Kuwelkar, 2017. *Real-time implementation of face recognition system*. International Conference on Computing Methodologies and Communication (ICCMC), pp. 249-255.

[8]. K. Koide, E. Menegatti, M. Carraro, M. Munaro, J. Miura, 2017. *People tracking and re-identification by face recognition for RGB-D camera networks*. European Conference on Mobile Robots (ECMR), pp. 1-7.

[9]. X. Tian, 2009. *Face Recognition System and It's Application*. First International Conference on Information Science and Engineering, Nanjing, pp. 1244-1245.

[10]. D. A. Chowdhry, A. Hussain, M. Z. Ur Rehman, F. Ahmad, A. Ahmad, M. Pervaiz, 2013. *Smart security system for sensitive area using face recognition*. CSUDET, Selangor, pp. 11-14.

[11]. Z. Jian, S. Wan-juan, 2010. *Face detection for security surveillance system.* 5th International Conference on Computer Science & Education, pp. 1735-1738.

[12]. P. A. Viola, M. J. Jones, 2001. *Rapid object detection using a boosted cascade of simple features*. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, I–I.

[13]. https://store.arduino.cc/usa/mega-2560-r3

[14]. https://www.arduino.cc/en/Guide/ArduinoEthernetShield/

[15]. W. Cort, M. Kenji, 2007. *On the use of dimensioned measures of error to evaluate the performance of spatial interpolators*. International Journal of Geographical Information Science. Vol.20, pp. 89–102.

[16]. http://nlprweb.ia.ac.cn/english/irds/facedatabase.htm

[17]. Florian Schroff, Dmitry Kalenichenko, James Philbin, 2015. *FaceNet: A Unified Embedding for Face Recognition and Clustering*. CoRR, Vol. abs/1503.03832, pp. 1-10.

[18]. Chunming Wu, Ying Zhang, 2021. *MTCNN and FACENET Based Access Control System for Face Detection and Recognition*. Aut. Control Comp. Sci. 55, pp. 102–112.

[19]. https://drive.google.com/drive/folders/1_ssN2plDuEFG2ydbnYkJ13X7MsPjq_cm

AUTHOR INFORMATION

Doan Thi Huong Giang

Faculty of Control and Automation, Electric Power University