

MỘT GIẢI PHÁP ĐIỂM DANH BẰNG CÔNG NGHỆ NHẬN DẠNG KHUÔN MẶT VÀ ỨNG DỤNG TẠI PHÒNG THỰC HÀNH KHOA CÔNG NGHỆ THÔNG TIN

Lương Minh Quân *, Nguyễn Tiến Hiền, Lê Văn Dũng, Lê Phương Thảo, Nguyễn Thị Huyền

Khoa Công nghệ thông tin, Học viện Nông nghiệp Việt Nam

**Tác giả chính liên hệ: lmquan@vnua.edu.vn*

Ngày nhận bài: 03.07.2023

Ngày chấp nhận đăng: 05.01.2024

TÓM TẮT

Mỗi gương mặt có một thuộc tính đơn nhất do đó hình ảnh gương mặt được sử dụng như một khóa bảo mật để truy cập tài khoản cá nhân trong các lĩnh vực như ngân hàng, dịch vụ thương mại điện tử, tài khoản cá nhân trên điện thoại hoặc máy tính. Nghiên cứu này nhằm mục đích đưa ra một giải pháp mới để điểm danh sinh viên tham dự lớp học bằng công nghệ nhận diện khuôn mặt khi kết hợp các mạng học sâu đa nhiệm để phát hiện khuôn mặt trong ảnh hoặc video, công nghệ mã hóa của mạng FaceNet để số hóa khuôn mặt phát hiện được và thuật toán phân lớp để tìm kiếm và so khớp khuôn mặt cần nhận diện với thông tin của khuôn mặt được lưu trữ trong cơ sở dữ liệu. Nghiên cứu đã thu được bộ cơ sở dữ liệu điểm danh sinh viên tham dự lớp học, chương trình máy tính thực hiện việc điểm danh bằng công nghệ nhận diện khuôn mặt và phân tích đánh giá hiệu quả của các chế độ điểm danh khác nhau bao gồm: sử dụng hình ảnh chụp của từng nhóm nhỏ sinh viên; sử dụng webcam kết nối máy tính theo các chế độ tự động và bán tự động. Với hệ thống điểm danh này, giáo viên có thể kiểm soát thông tin tham dự lớp học của sinh viên, dễ dàng phát hiện gian lận của sinh viên trong học tập tại các lớp học phân (lý thuyết và thực hành) và thi hết học phần của sinh viên.

Từ khóa: Mạng học sâu đa nhiệm, FaceNet, thuật toán phân lớp, phát hiện khuôn mặt, nhận dạng khuôn mặt.

A Solution for Attendance Checking with Face Recognition Technology and Application at Laboratories in Faculty of Information Technology

ABSTRACT

Each human face has a unique characteristic so the face image is used as a security key to access personal accounts in many domains such as bankings, E-commerce services, phone or computer personal accounts. This study aimed to introduce a new solution for class attendance checking by using face recognition technology when combining multi-tasking deep learning networks, to detect faces in photos or videos, encryption of FaceNet to digitize detected faces and support vector machine clustering algorithm to search and match the face to be recognized with the face stored in the database. This research has obtained a database of student attendance, a computer program that implements the attendance by facial recognition technology, and analyzed and evaluated the effectiveness of other attendance methods, included: using photographs of small group of student and using a webcam connected to a computer in automatic and semi-automatic modes. With this kind of attendance checking system, teachers can control students' class attendance, easily detect cheating in study and examination in both theory and practicum courses and the final exams.

Keywords: Multi-tasking deep learning networks, FaceNet, support vector machine, face detection, face recognition, class attendance.

1. ĐẶT VẤN ĐỀ

Thái độ học tập, tham dự lớp học cũng như gian lận trong học tập và thi cử đang là vấn đề

nan giải đối với công tác quản lý học tập của sinh viên.

Với những hình thức điểm danh thông thường đang được áp dụng hiện nay như: gọi tên

từng sinh viên; điểm danh theo thứ tự vị trí ngồi; làm bài kiểm tra 5 phút để lấy thông tin sinh viên tham dự lớp; đếm số đầu người trong lớp;... Các cách thức điểm danh ở trên chỉ cho phép giảng viên nắm được thông tin tham dự lớp học của sinh viên một cách tương đối và gần như không thể để phát hiện được tình trạng gian lận với trường hợp sinh viên thuê một sinh viên khác đi học và điểm danh thay cho mình.

Vấn đề trên đòi hỏi nhà quản lý và giáo viên giảng dạy lớp học phần cần có một phương pháp tiếp cận mới trong việc quản lý học tập và thi cử nhằm phát hiện và ngăn chặn các gian lận xuất hiện trong lớp học, kịp thời đưa ra cảnh báo tới sinh viên và gia đình sinh viên biết được thực trạng học tập hiện tại của con em mình.

Nghiên cứu về công nghệ nhận dạng bằng khuôn mặt được phát triển và ứng dụng trong nhiều lĩnh vực khác nhau như: Ứng dụng công nghệ nhận dạng khuôn mặt trong quản lý hành chính của nhóm sinh viên trường Đại học Thái Nguyên (2017). Yugashini & cs. (2013), đã phát triển mô hình của tự động đóng mở khi kết hợp thông tin nhận dạng khuôn mặt và bộ cảm biến

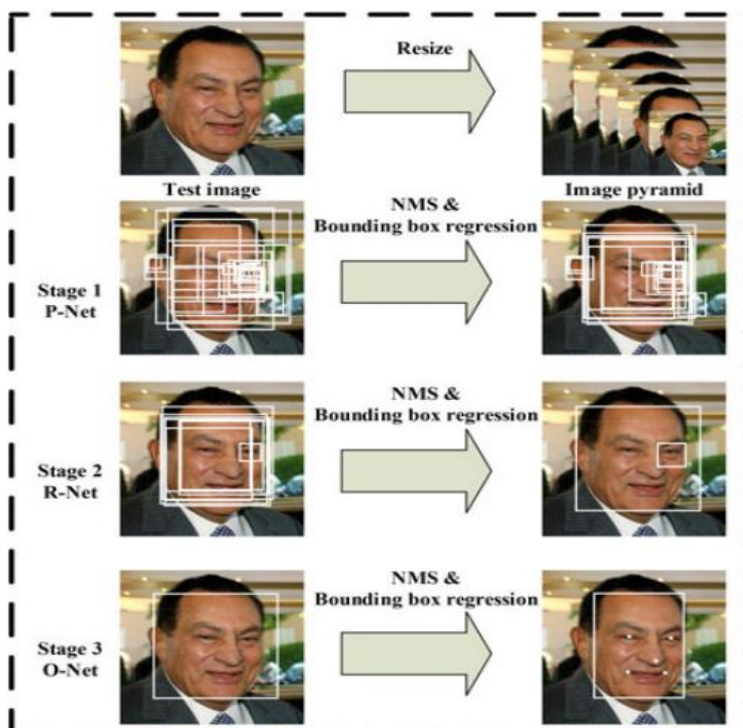
PIC Microcontroller 16F877A. Ngô Tùng Sơn & cs. (2020), đã nghiên cứu và phát triển hệ thống hỗ trợ điểm danh với camera giám sát tại giảng đường sử dụng mạng học sâu (DCNN) tại Đại học FPT. Trần Như Ý & Nguyễn Phương Hạc (2021), đã kết hợp Haar-Cascade và mạng học sâu đa nhiệm (MTCNN) vào thuật toán ARCFACE (Additive Angular Margin Loss) giúp cải tiến về mặt tốc độ xử lý và giảm thời gian nhận dạng khuôn mặt.

Trong nghiên cứu này chúng tôi sử dụng công nghệ nhận diện của DeepFace được phát triển bởi FaceBook để phát triển công nghệ lõi của chương trình nhận diện

2. PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Mạng học sâu đa nhiệm MTCNN

MTCNN là một mạng nơ-ron có thể thực hiện nhận diện khuôn mặt và căn chỉnh khuôn mặt trên hình ảnh. Zhang & cs. (2016) đã giới thiệu mạng nơ-ron này bao gồm 3 mạng riêng biệt (Hình 1).



Nguồn: Zhang & cs. (2016).

Hình 1. Các giai đoạn của mạng đa nhiệm MTCNN

2.1.1. Mạng P (Proposal Network)

Một bộ hình ảnh, là bản sao của ảnh gốc với kích thước thay đổi theo cấu trúc kim tự tháp (Image Pyramid) được đưa qua mạng P để xác định các hộp giới hạn chứa khuôn mặt và giá trị mức độ tin cậy của các hộp này. Thuật toán loại trừ để loại bỏ các hộp giới hạn có độ tin cậy nhỏ hơn ngưỡng và thuật toán NMS (No - Max Supression) để loại bỏ các hộp giới hạn có tỉ lệ trùng nhau vượt ngưỡng cho phép.

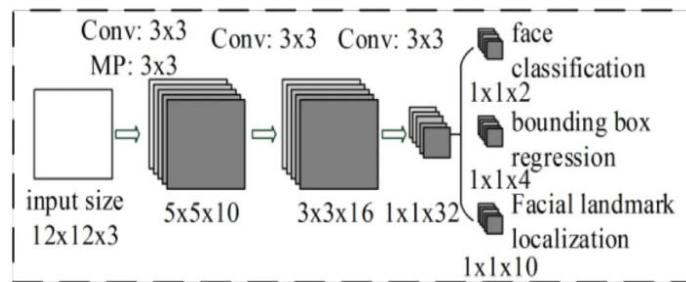
2.1.2. Mạng R (Refine Network)

Mạng R (Refine Network) thực hiện các bước như mạng P-Net. Tuy nhiên, mạng còn sử dụng một phương pháp có tên là zero-padding, nhằm thực hiện việc chèn thêm các điểm ảnh có giá trị 0 vào các phần thiếu của hộp giới hạn nếu hộp giới hạn này nằm ngoài biên của ảnh. Tất cả các hộp giới hạn này sẽ được điều chỉnh

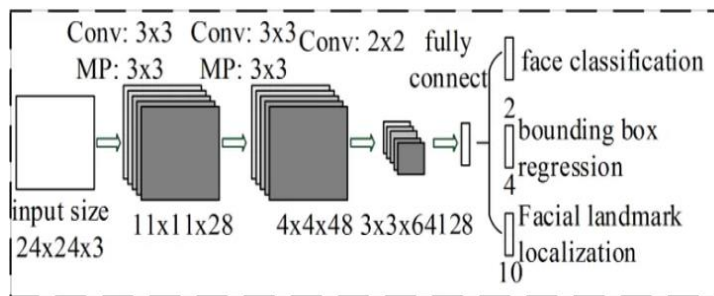
về kích thước 24×24 và được truyền vào mạng R-Net. Kết quả đầu ra của mạng R-Net cho biết đầu vào có phải là khuôn mặt hay không, bao gồm 1 véctơ 4 phần tử mô tả biên giới hạn của khuôn mặt và 1 véctơ có 10 phần tử mô tả vị trí trên khuôn mặt đó. Đầu ra của mạng này sẽ được truyền vào mạng tiếp theo, mạng O-Net.

2.1.3. Mạng O (Output Network)

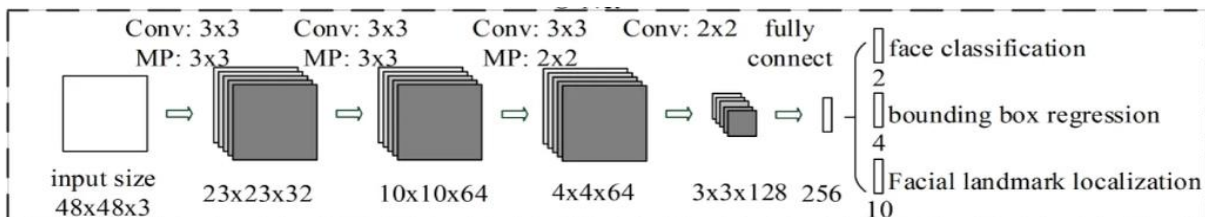
Mạng này cũng thực hiện chức năng tương tự như trong mạng R-Net, với kích thước ảnh ở đầu vào của mạng được thay đổi thành 48×48 . Tuy nhiên, kết quả đầu ra của mạng lúc này không còn chỉ là các tọa độ của các hộp giới hạn nữa. Tương ứng với mỗi khuôn mặt nhận dạng được trong ảnh, kết quả trả về 3 bộ giá trị bao gồm: 4 tọa độ của hộp giới hạn; 5 tọa độ điểm cơ sở trên khuôn mặt mặt bao gồm: 2 mắt, 1 mũi, 2 bên cánh môi và hệ số độ tin cậy của mỗi hộp giới hạn.



Hình 2. Mạng đề xuất P



Hình 3. Mạng tinh chỉnh R



Hình 4. Mạng O

2.1.4. Dữ liệu huấn luyện của mạng MTCNN

Zhang & cs. (2016) đã tiến hành so sánh hiệu quả của mạng MTCNN trong phát hiện khuôn mặt với các phương pháp tân tiến nhất trên các tập dữ liệu ảnh lớn như FDDB (Face Detection Data Set and Benchmark), WIDER FACE, AFLW (Annotated Facial Landmarks in the Wild). Thử nghiệm cho kết quả vượt trội về độ chính xác và duy trì được tốc độ xử lý theo thời gian thực.

Kiến trúc mạng học sâu đa nhiệm MTCNN tương đối phức tạp để triển khai một cách độc lập trên tập dữ liệu mới. Do đó, trong bài báo này nhóm nghiên cứu sử dụng các mô hình được huấn luyện từ trước trên các tập dữ liệu lớn bởi Ivan de Paz Centeno và được cấp phép bởi học viện công nghệ Massachusetts (MIT) dưới dạng mã nguồn mở.

2.2. Mạng FaceNet

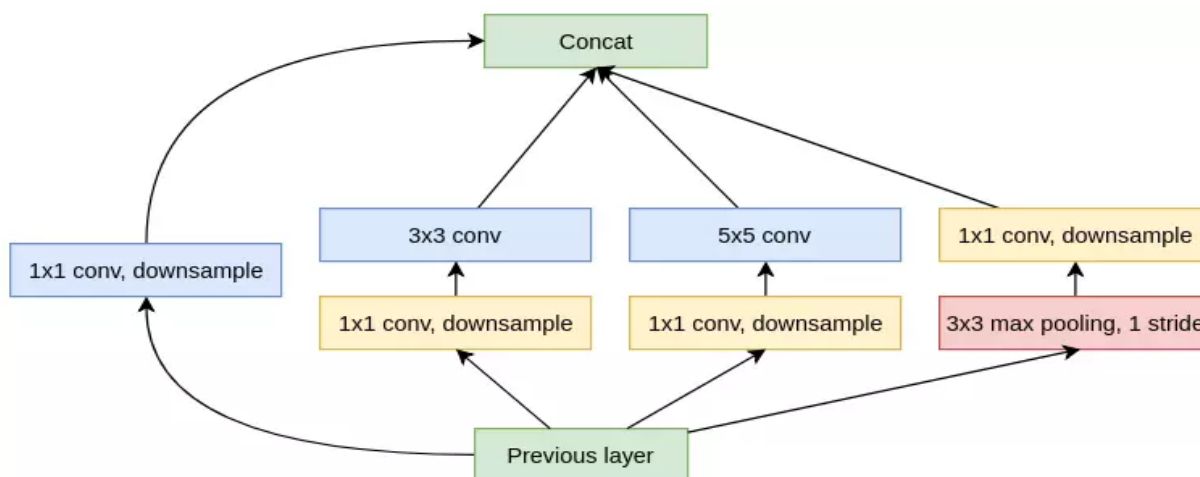
2.2.1. Vectơ nhúng

Là một vectơ với số chiều cố định, giá trị của vectơ này đã được học trong quá trình huấn

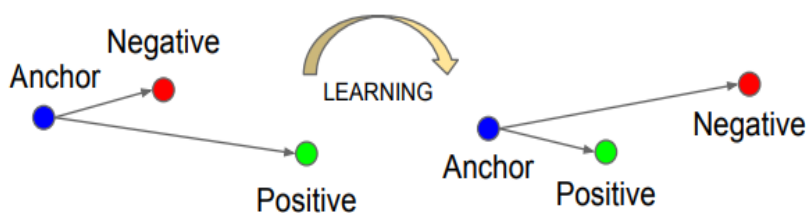
luyện của mô hình và đại diện cho một tập hợp các đặc trưng của khuôn mặt, đóng vai trò quan trọng trong việc phân loại các đối tượng trong chiều không gian đã được biến đổi. Vectơ nhúng rất hữu dụng trong việc tìm các lân cận trong một cụm cho sẵn, dựa vào mối quan hệ theo khoảng cách giữa các vectơ nhúng với nhau. Trong nghiên cứu về số chiều số hóa của vectơ nhúng, Schroff. & cs. (2015) đã chỉ ra rằng vectơ nhúng 128-D cho kết quả tối ưu về độ chính xác.

2.2.2. Mạng Inception

Năm 2014, Google đã giới thiệu một cấu trúc mạng CNN (Hình 5), với đặc trưng là các khối Inception cho phép mạng được học theo cấu trúc song song, nghĩa là với 1 đầu vào có thể được đưa vào nhiều lớp tích chập khác nhau để đưa ra các kết quả khác nhau, sau đó được gộp lại thành 1 kết quả duy nhất ở đầu ra. Việc học song song này giúp mạng có thể học được nhiều chi tiết hơn, lấy được nhiều đặc trưng hơn so với mạng CNN truyền thống. Ngoài ra, mạng cũng áp dụng các khối tích chập có kernel bằng 1×1 nhằm giảm kích thước của mạng, khiến việc huấn luyện trở nên nhanh hơn.



Hình 5. Cấu trúc mạng Inception V1



Hình 6. Cấu trúc của Triplet Loss

2.2.3. Hàm mất mát 3 điểm (Triplet Loss)

Trong mạng CNN với hàm Loss truyền thống, ở đó ta chỉ so sánh giá trị đầu ra của mạng với giá trị thực (ground truth) của dữ liệu. Schroff. & cs. (2015) đã đưa ra hàm mất mát 3 điểm (triplet loss) bao gồm 3 giá trị đầu vào: mô neo x_i^a (anchor) - là ảnh khuôn mặt đầu ra của mạng; x_i^p (positive) - là ảnh khuôn mặt của cùng một người giống với anchor; x_i^n (negative) - là ảnh khuôn mặt của một người khác với anchor. Yêu cầu của thuật toán Triplet Loss là cực tiểu hóa khoảng cách euclid (Euclidean distance) giữa x_i^a và x_i^p (anchor - positive) tương ứng với ảnh khuôn mặt của cùng một nhân dạng và cực đại hóa khoảng cách giữa x_i^a và x_i^n (anchor - negative) tương ứng với ảnh khuôn mặt của các nhân dạng khác nhau (Hình 6).

$$\|f(x_i^a) - f(x_i^p)\|_2^2 + \alpha < \|f(x_i^a) - f(x_i^n)\|_2^2 \quad (1)$$

Trong đó:

- $f(x_i^a), f(x_i^p), f(x_i^n)$ là các vectơ nhúng mã hóa của ảnh x_i^a, x_i^p, x_i^n

- $f(x_i^a), f(x_i^p), f(x_i^n) \in \Gamma$ là tập hợp các bộ ba triplets trong bộ dữ liệu đào tạo có bậc là N

- α (margin) là độ sai lệch tối thiểu giữa $f(x_i^p)$ và $f(x_i^n)$

Để mạng học được nhiều “tri thức” hơn, trong mỗi bó dữ liệu (mini-batch) điểm positive được chọn nằm xa nhất có thể so với anchor, điểm negative được chọn nằm gần nhất có thể so với anchor, các hàm argmin và argmax được sử dụng để tìm ra các cặp triplets thỏa mã yêu cầu về khoảng cách.

Hàm mất mát (Loss) tổng quát được xác định bởi công thức sau:

$$L = \sum_{i=1}^N \left[\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right] \quad (2)$$

2.2.4. Mô hình pre-trained với FaceNet

Trong nghiên cứu này, nhóm sử dụng mô hình đã được đào tạo từ trước (pre-trained

model) của mạng học sâu FaceNet được nghiên cứu bởi Schroff & cs. (2015) khi sử dụng mạng nơ-ron tích chập sâu (DNN) với cấu trúc các Inception của Szegedy & cs. (2015) trên tập dữ liệu LFW (Labeled Faces in the Wild) và DB (Youtube Faces) để nhận dạng khuôn mặt.

2.3 Kết hợp mạng MTCNN và FaceNet

Trong nghiên cứu này, nhóm nghiên cứu sử dụng mạng học sâu đa nhiệm MTCNN để phát hiện khuôn mặt người trong bức ảnh cần phân tích. Mỗi hình ảnh khuôn mặt phát hiện được từ mạng này sẽ được cắt và thay đổi kích thước thành kích thước chuẩn 160×160 pixels và được truyền qua mạng FaceNet để số hóa thành vectơ 128 chiều. Vectơ số hóa này được lưu lại thành bộ cơ sở dữ liệu cơ bản và được sử dụng để so sánh mức độ tương tự nhau giữa các khuôn mặt bằng cách tích khoảng cách giữa các vectơ với nhau thông qua thuật toán phân lớp SVM.

3. KẾT QUẢ VÀ THẢO LUẬN

Để hoàn thành việc thu thập dữ liệu và phát triển chương trình nhận diện dựa trên mạng học sâu đa nhiệm MTCNN, FaceNet và thuật toán phân lớp SVM, nhóm nghiên cứu đã tiến hành thu thập dữ liệu ảnh khuôn mặt của sinh viên các nhóm thực hành tại Bộ môn Vật lý, Khoa Công nghệ thông tin trong khoảng thời gian từ tháng 1/2023 tới tháng 6/2023. Giao diện phần mềm được thiết kế bởi PyQt5 Designer, dựa trên ngôn ngữ lập trình Python.

3.1. Thu thập dữ liệu ảnh sinh viên tại Phòng thực tập Vật lý, Khoa Công nghệ thông tin

Quá trình thu thập dữ liệu ảnh của 200 sinh viên được thực hiện bởi phần mềm chụp ảnh tự động, được phát triển bởi nhóm nghiên cứu. Số lượng ảnh chụp với mỗi mẫu là 30 ảnh. Để việc phân tích dữ liệu và nhận dạng khuôn mặt được chính xác thì quá trình chụp ảnh khuôn mặt phải thỏa mãn một số điều kiện sau:

- Người được lấy mẫu không đeo khẩu trang, cách camera từ 40 đến 50cm để thu được

hình ảnh khuôn mặt rõ nét, không bị mờ trong khung hình kích thước 640×480 pixels.

- Người được lấy mẫu được khuyến nghị có tương tác với nhóm nghiên cứu trong quá trình chụp hình, để thu được ảnh chụp với nhiều trạng thái cảm xúc khác nhau (cười, há miệng, nháy mắt, nhắm mắt,...), để phong phú dữ liệu số hóa và nâng cao độ chính xác trong quá trình nhận dạng.

- Người được lấy mẫu sẽ thay đổi vị trí góc của khuôn mặt so với camera để thu được các góc nhìn khác nhau: quay đầu từ bên trái (-45°) tới bên phải ($+45^\circ$); quay đầu từ góc nhìn chính diện hướng lên cao 10° . Thời gian lấy mẫu là 15s.

- Khi chụp ảnh, có thể thay đổi mức độ chiếu sáng lên khuôn mặt để thu được các góc phối sáng khác nhau, điều này sẽ có đóng góp rất quan trọng khi nhận dạng ở các điều kiện ánh sáng khác nhau của môi trường lớp học, giảng đường, hội trường,...

3.2. Phân tích số liệu số hóa khuôn mặt

3.2.1. Phương pháp phân tích thành phần chính PCA

Phương pháp phân tích thành phần chính (PCA) được sử dụng để phân tích sự phân bố trong không gian 3 chiều cấu trúc dữ liệu của véctơ nhúng 128-D cho tập mẫu 200 bộ dữ liệu. Các đồ thị trên hình 7, biểu diễn dữ liệu của véctơ nhúng 128-D được mã hóa bởi mạng FaceNet với số lượng ảnh chứa khuôn mặt $N_{\text{imgs}/\text{person}}$ thay đổi từ 1 đến 30. Điểm dữ liệu được phân bố nằm trong hình cầu bán kính R_0 . Sự phân bố của các điểm dữ liệu này tương đối hỗn loạn và chưa có tính tập trung với $N_{\text{imgs}/\text{person}} < 5$. Khi $N_{\text{imgs}/\text{person}} \geq 10$, các điểm ảnh của mỗi bộ dữ liệu khác có xu hướng tập trung theo từng cụm, khoảng cách giữa các cụm này trở nên bé hơn. Với $N_{\text{imgs}/\text{person}} \geq 20$, bộ 200 dữ liệu được phân bố tương đối ổn định trong hình cầu có bán kính $r \sim \frac{1}{7}R_0$.

Như vậy việc lựa chọn số lượng ảnh của mỗi nhân dạng $N_{\text{imgs}/\text{person}}$ trong bộ dữ liệu mã

hóa có thể được đánh giá một cách định tính dựa vào phân tích ở trên. Trong nghiên cứu này, chúng tôi sử dụng $N_{\text{imgs}/\text{person}} = 30$ để cân đối giữa độ chính xác của kết quả nhận dạng và thời gian xử lý cũng như tài nguyên của máy tính hiện có.

3.2.2. Phương pháp phân tích dựa trên các thuật toán máy học

Để đánh giá bộ dữ liệu kiểm tra của 200 (mẫu), mỗi mẫu có 30 ảnh, mỗi ảnh được mã hóa thành véctơ 128 - D, trong đó 80% dữ liệu cho tập đào tạo và 20% dữ liệu cho tập kiểm tra trong bảng 1.

Nhóm nghiên cứu sử dụng các mô hình học máy từ thư viện sklearn để đánh giá độ chính xác của các thuật toán. Kết quả được mô tả trong bảng 2. Từ kết quả trên cho thấy, các mô hình học máy KNN và SVM cho kết quả có độ chính xác tới trên 99% với bộ dữ liệu được số hóa bởi mạng MTCNN - FaceNet.

Từ những phân tích ở trên, thuật toán phân lớp SVM với vai trò là bước thứ ba của quá trình nhận diện để xác định thông tin của khuôn mặt cần nhận diện.

3.3. Triển khai ứng dụng tại Phòng Thí nghiệm Vật lý, Khoa Công nghệ thông tin, Học viện Nông nghiệp Việt Nam

3.3.1. Các chế độ điểm danh

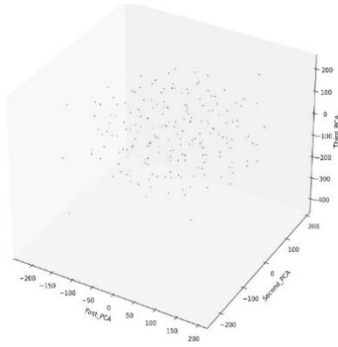
Phương pháp điểm danh tham dự lớp học bằng công nghệ nhận diện được nhóm nghiên cứu triển khai tại các nhóm thực hành vật lý, Khoa Công nghệ thông tin. Do tính chất riêng biệt của học phần thực hành vật lý: mỗi nhóm thực hành có tối đa 25 sinh viên được chia thành 7 nhóm nhỏ, mỗi nhóm nhỏ có từ 3 đến 4 sinh viên cùng hợp tác với nhau để làm bài thực hành. Do đó, nhóm nghiên cứu đã tiến hành điểm danh bằng công nghệ nhận diện trong một số trường hợp sau:

- Nhận diện thông qua ảnh chụp các nhóm sinh viên trong lớp học: yêu cầu với ảnh chụp cần phải rõ nét, các khuôn mặt không bị che khuất, góc chụp trực diện.

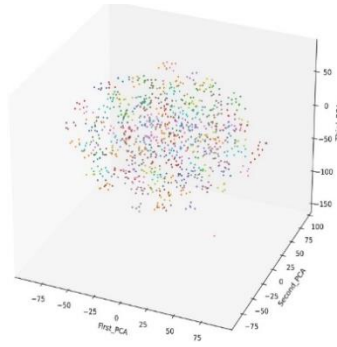
Một giải pháp điểm danh bằng công nghệ nhận dạng khuôn mặt và ứng dụng tại Phòng Thực hành Khoa Công nghệ thông tin

- Nhận diện theo thời gian thực thông qua hệ thống camera giám sát hoặc webcam: việc điểm danh theo thời gian thực được thực hiện khi mỗi sinh viên hoặc từng nhóm sinh viên đứng

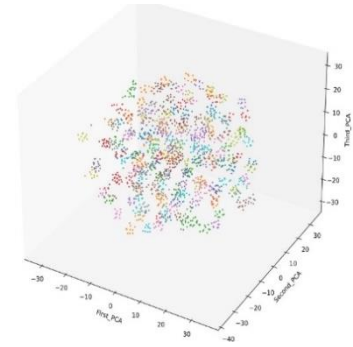
trước camera, hệ thống sẽ phân tích và nhận diện từng người trong mỗi khung hình. Nhóm đã nghiên cứu và phát triển công nghệ nhận diện theo thời gian thực với 3 chế độ cơ bản:



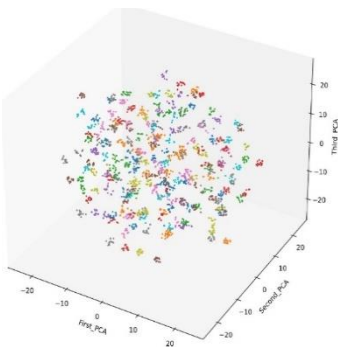
Mỗi dữ liệu chứa 1 ảnh



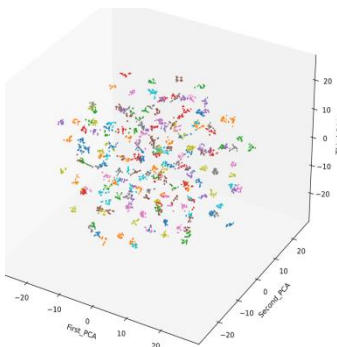
Mỗi dữ liệu chứa 5 ảnh



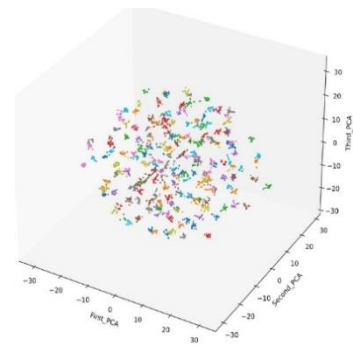
Mỗi dữ liệu chứa 10 ảnh



Mỗi dữ liệu chứa 15 ảnh



Mỗi dữ liệu chứa 20 ảnh



Mỗi dữ liệu chứa 30 ảnh

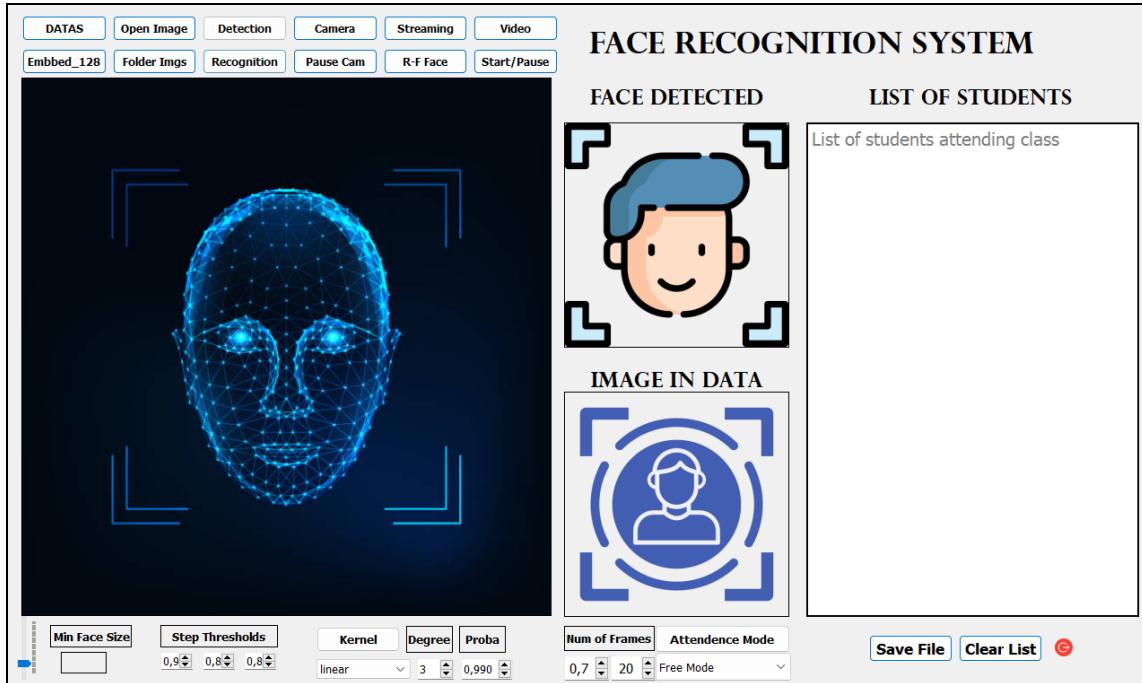
Hình 7. Phân bố không gian của điểm ảnh được mã hóa bởi FaceNet

Bảng 1. Tham số bộ dữ liệu số hóa của 200 sinh viên

Số thứ tự	Số mẫu	Số ảnh	Số chiều mã hóa
Tập dữ liệu	200	6000	128
Tập training	160	4800	128
Tập testing	40	1200	128

Bảng 2. So sánh kết quả các mô hình máy học

Phương pháp máy học	Tham số	Độ chính xác
Multilinear Regression	Mặc định	49,43%
Decision Tree Regressor	Mặc định	58,17%
Polynomial with Ridge	Degree = 3, alpha = 1	83,15%
Polynomial Regression	Degree = 3	85,08%
Random Forest Regressor	Mặc định	89,23%
K nearest neighbor (KNN)	N_neighbors = 4	99,84%
Support Vector Machine (SVM)	Kernel = 'linear' C = 1	100%



Hình 8. Giao diện chính của phần mềm nhận dạng khuôn mặt



Hình 9. So khớp hình ảnh (hình ảnh tác giả thực hiện nghiên cứu)

+ *Chế độ tự do (Free mode)*: trong chế độ này, thông tin cá nhân về mỗi khuôn mặt trong khung hình sẽ được thông báo, khi nhìn vào màn hình thì giáo viên có thể biết sinh viên nào (tên, MSV,...)

+ *Chế độ điểm danh có kiểm soát*: chế độ này cho phép giáo viên có thể kiểm soát được thông tin nhận diện bởi phần mềm có đúng với dữ liệu sinh viên được lưu trữ trong cơ sở dữ liệu hay không. Để thực hiện ý tưởng nêu trên, mỗi khi hình ảnh khuôn mặt trong khung hình được phát hiện, hệ thống sẽ so khớp khuôn mặt đó với khuôn mặt được lưu trữ trên cơ sở dữ liệu thông qua cơ chế số hóa của vectơ nhúng để tìm ra vectơ nhúng có khoảng cách gần nhất tương ứng với khuôn mặt cần nhận dạng. Khi đó, một bảng thông báo được gửi tới người quản lý với thông

điệp: “2 khuôn mặt này có phải cùng một nhân dạng hay không?”. Việc so khớp hình ảnh bằng mắt thường có thể được thực hiện một cách dễ dàng, do đó việc điểm danh sinh viên có độ chính xác cao và tránh được các trường hợp gian lận của người học như: thuê người đi học, thuê người đi thi, phát hiện người lạ mặt (không phải thành viên của lớp),...

+ *Chế độ điểm danh không kiểm soát*: trong chế độ này, việc điểm danh được thực hiện một cách tự động mà không có sự can thiệp từ phía giáo viên hay người quản lý phần mềm. Để thông tin nhận dạng đảm bảo độ chính xác cao, nhóm nghiên cứu đã bổ sung thuật toán nhận dạng trên nhiều khung hình liên tiếp với số lượng N từ 20 đến 30 khung hình (frame), do vậy mỗi sinh viên sẽ đứng trước camera trong

Một giải pháp điểm danh bằng công nghệ nhận dạng khuôn mặt và ứng dụng tại Phòng Thực hành Khoa Công nghệ thông tin

thời gian 2-3 giây để hoàn tất quá trình phân tích. Vì mỗi khung hình sẽ cho kết quả nhận dạng của một người nào đó trong cơ sở dữ liệu, nên 20 khung hình sẽ cho 20 kết quả có thể giống hoặc khác nhau. Kết quả nhận dạng được tự động lưu và cơ sở dữ liệu nếu số lượng khung hình trả về cùng một người và phải thỏa mãn điều kiện:

$$\text{Kết quả nhận dạng} \geq \text{round}(\text{coeff} * N) \quad (3)$$

Trong đó, coeff là hệ số trong khoảng [0, 1]; N là số lượng khung hình cần phân tích. Các giá trị này được điều chỉnh dựa trên các điều kiện thực tế của quá trình nhận dạng như: độ phổ

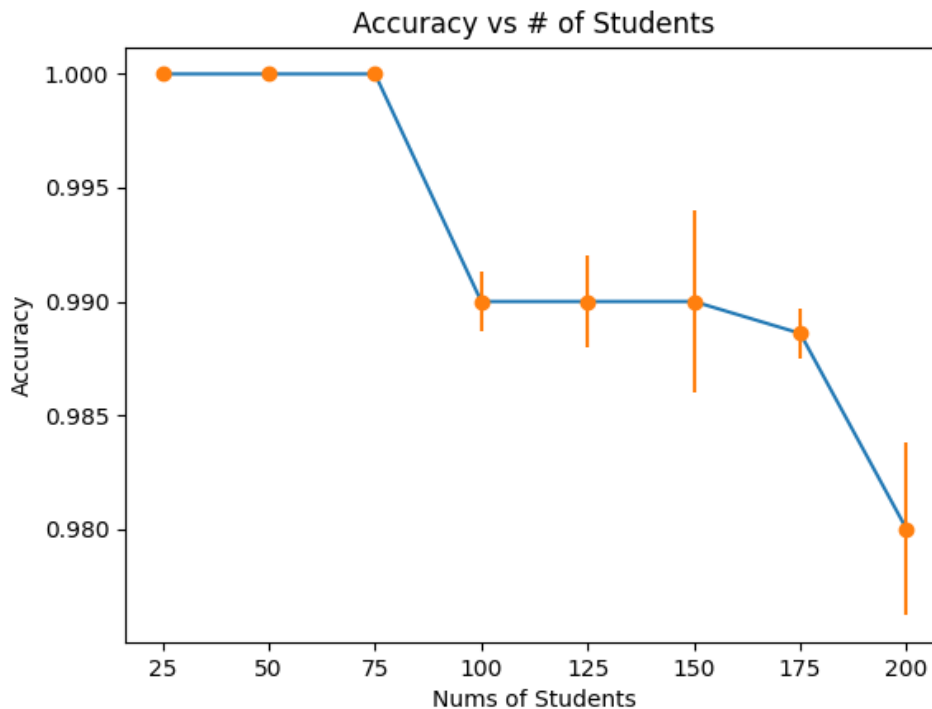
biến của hình ảnh khuôn mặt; điều kiện chiếu sáng; số lượng sinh viên trong bộ cơ sở dữ liệu,...

3.3.2. Kết quả nhận diện qua ảnh chụp từng nhóm sinh viên thực hành

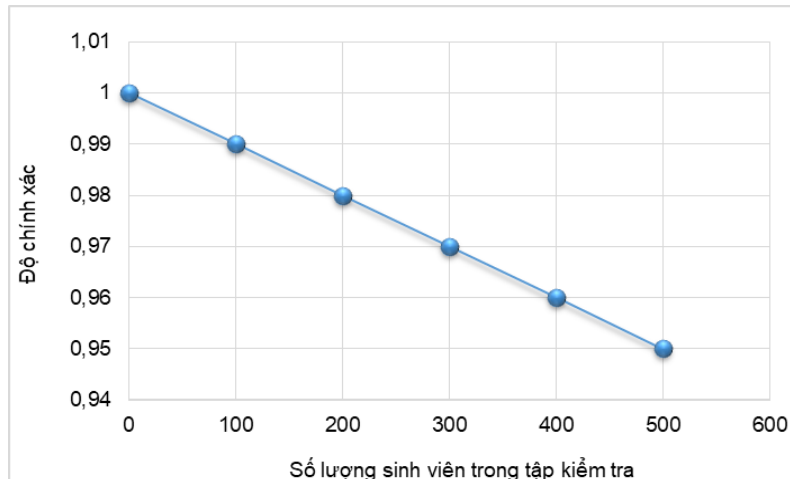
Ảnh chụp của từng nhóm nhỏ được tiến hành tiền xử lý trước khi đưa vào phần mềm nhận diện để đảm bảo kết quả có độ chính xác trên các tập cơ bản có 25 sinh viên là 100%. Từ bảng số liệu 3 và hình 10 ta nhận thấy độ chính xác có xu hướng giảm dần khi số lượng mẫu trong bộ dữ liệu cần nhận diện tăng lên cứ trung bình 1% cho 100 mẫu và được ngoại suy theo hình 11 khi kích thước mẫu tăng lên tới 500 sinh viên.

Bảng 3. Phân tích kết quả nhận diện theo ảnh chụp

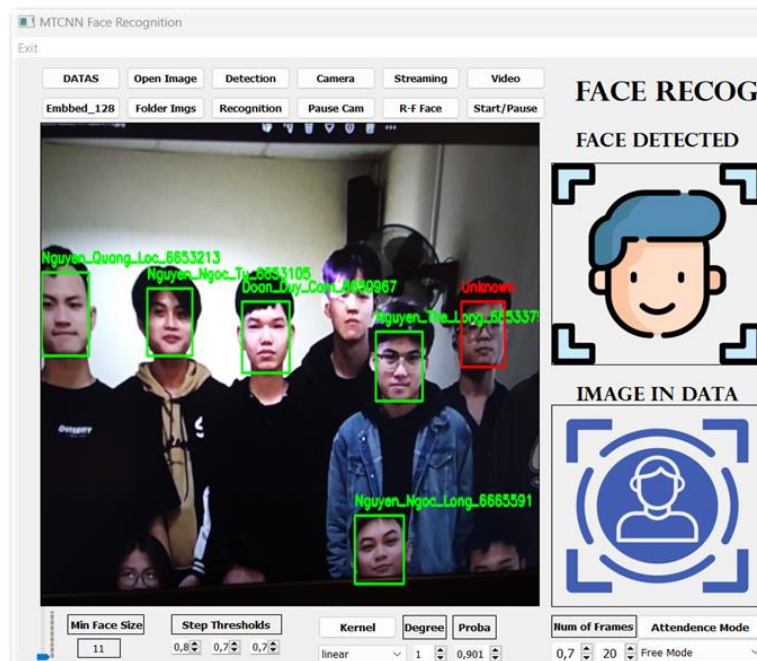
Số lượng sinh viên	Độ chính xác	Sai số
25	1,0000	0,0000
50	1,0000	0,0000
75	1,0000	0,0000
100	0,9900	0,0013
125	0,9900	0,0020
150	0,9900	0,0040
175	0,9886	0,0011
200	0,9800	0,0038



Hình 10. Phân tích kết quả nhận diện theo số lượng sinh viên trong mẫu



Hình 11. Ngoại suy mức độ suy giảm của độ chính xác khi kích thước mẫu tăng lên



Hình 12. Chương trình nhận dạng trong chế độ tự do (Free Mode)

Phân tích trên cho thấy rằng, khi triển khai hệ thống nhận dạng qua ảnh chụp với tập dữ liệu có kích thước lớn, chúng ta cần phải thực hiện thêm một số phân tích bổ xung như phân chia tập dữ liệu theo giới tính; phân chia theo độ tuổi;... để chia nhỏ tập dữ liệu thành các tập có kích thước nhỏ hơn 200.

3.3.3. Kết quả nhận dạng theo thời gian thực qua hệ thống camera giám sát tự động

Để tiến hành điểm danh sinh viên tham dự lớp học thông qua hệ thống camera có kết nối

với máy tính, một camera được kết nối với máy tính, ở độ cao 170cm (tương đương độ cao trung bình của sinh viên) được bố trí tại cửa ra vào của lớp học. Khi sinh viên vào lớp, lần lượt từng người hoặc từng nhóm từ 3 đến 4 người sẽ đứng trước camera, mặt nhìn thẳng hướng về phía camera trong khoảng thời gian từ 1 đến 2 giây để hệ thống thu lại hình ảnh và xử lý, kết quả nhận dạng sẽ hiển thị trên cửa sổ màn hình của máy tính. Nhóm nghiên cứu đã phát triển 3 chế độ hoạt động khác nhau của chương trình nhận dạng.

Một giải pháp điểm danh bằng công nghệ nhận dạng khuôn mặt và ứng dụng tại Phòng Thực hành Khoa Công nghệ thông tin

a. Chế độ tự do (Free Mode)

Trong chế độ này, từng nhóm sinh viên sẽ đứng trước camera ở khoảng cách nhỏ hơn 1m. Camera sẽ thu lại hình ảnh với tốc độ 30 khung hình (frame)/giây. Mỗi khung hình sẽ được phân tích bởi chương trình để phát hiện các khuôn mặt. Với mỗi khuôn mặt phát hiện được sẽ được mã hóa bởi mạng FaceNet. Cuối cùng thuật toán phân lớp SVM (Support Vector Machine) được sử dụng để so khớp khuôn mặt phát hiện được với nhân dạng trong bộ dữ liệu cơ sở.

Như đã biết, mỗi nhân dạng được phát hiện bởi hệ thống sẽ đi kèm theo một hệ số γ mô tả độ chính xác của kết quả nhận dạng. Hệ số này sẽ được so sánh với giá trị ngưỡng được điều chỉnh bởi người dùng ($Proba = 0,901$). Nếu $\gamma \geq Proba$ thông tin sinh viên cần nhận dạng sẽ được hiển thị. Nếu $\gamma < Proba$, một khung hình mẫu đồ kết hợp với dòng chữ “Unknown” mô tả trạng thái chưa xác định được danh tính của sinh viên đó. Với cấu hình của máy tính Core i7 - 12700, card đồ họa NVIDIA GeForce RTX 3070 Ti Laptop GPU, chương trình có khả năng nhận dạng đồng thời 10 khuôn mặt trong khi vẫn đảm bảo tốc độ 30 khung hình /giây (FPS = 30).

b. Hình thức điểm danh tự động có can thiệp (Authorize Checking)

Khi chương trình chạy trong chế độ Authorize Checking (Hình 13), giáo viên yêu cầu từng sinh viên đứng trước camera để ghi lại hình ảnh. Hệ thống phần mềm sẽ phân tích thông tin trên từng khung hình và đưa ra thông báo về kết quả nhận dạng nếu xuất hiện khuôn mặt trong khung hình đó. Một thông báo xuất hiện với câu hỏi “Are you Ho_Va_Ten_MSV?”, kèm theo cặp hình ảnh của khuôn mặt xuất hiện trong cửa sổ FACE DETECTED và hình ảnh đối chiếu trong cửa sổ IMAGE IN DATA.

Giáo viên hoặc người quản lý sẽ đối chiếu hình ảnh nằm trong 2 cửa sổ này để xác nhận thông tin nhận dạng có chính xác hay không:

+ Chọn “Yes” nếu 2 khuôn mặt trong FACE DETECTED và IMAGE IN DATA trùng nhau. Khi đó thông tin cá nhân của người được nhận dạng sẽ xuất hiện trong danh sách “LIST OF STUDENTS” và được lưu vào danh sách điểm danh trong cơ sở dữ liệu.

+ Chọn “No” nếu hai hình ảnh này không mô tả thông tin của cùng một nhân dạng, giáo viên có thể yêu cầu sinh viên thay đổi vị trí, khoảng cách so với camera hay thậm chí là trạng thái cảm xúc trên khuôn mặt,... để thu được kết quả nhận dạng giống với hình ảnh cần phân tích. Khi kết thúc, giáo viên có thể kiểm tra thông tin điểm danh bằng cách truy cập vào cơ sở dữ liệu “Student_HVN_Database.db”. Bảng “HVN_Class_MMTA_K66” lưu trữ thông tin điểm danh của lớp học theo từng buổi học riêng biệt.

Khi truy cập vào thông tin như hình 14, thông tin tham dự lớp học của sinh viên với các trường thông tin như: Họ và tên; MSV; thời gian đến lớp; và đặc biệt hình ảnh khuôn mặt cần nhận dạng và hình ảnh lưu trữ trong cơ sở dữ liệu được tích hợp trong trường dữ liệu “Image_Composed” giúp ta phân biệt được kết quả nhận dạng là đúng hay sai. Với những sinh viên tham dự lớp học sẽ có dấu tích “X” trong trường “Attendance”.

Hình thức điểm danh có can thiệp của người quản lý phù hợp để triển khai trong các lớp học tại giảng đường, phòng thực hành, các buổi sinh hoạt chuyên môn, các buổi thi học phần,...

c. Hình thức điểm danh tự động không can thiệp (Self-Checking)

Với hình thức điểm danh này, quy trình thực hiện tương tự như quá trình điểm danh tự động có can thiệp. Tuy nhiên điểm khác biệt nằm ở chỗ:

+ Không cần người quản lý xác nhận “Yes” hoặc “No”

+ Hệ thống sẽ tự phân tích thông tin người cần nhận dạng có đúng với hình ảnh lưu trữ hay không thông qua việc bổ xung thêm ràng buộc số khung hình tối thiểu (*Num of frames*) mà hệ thống cần phân tích phải thuộc về cùng một nhân dạng. Nếu yêu cầu số khung hình tối thiểu được thỏa mãn, hệ thống sẽ tự động lưu trữ thông tin vào bảng cơ sở dữ liệu

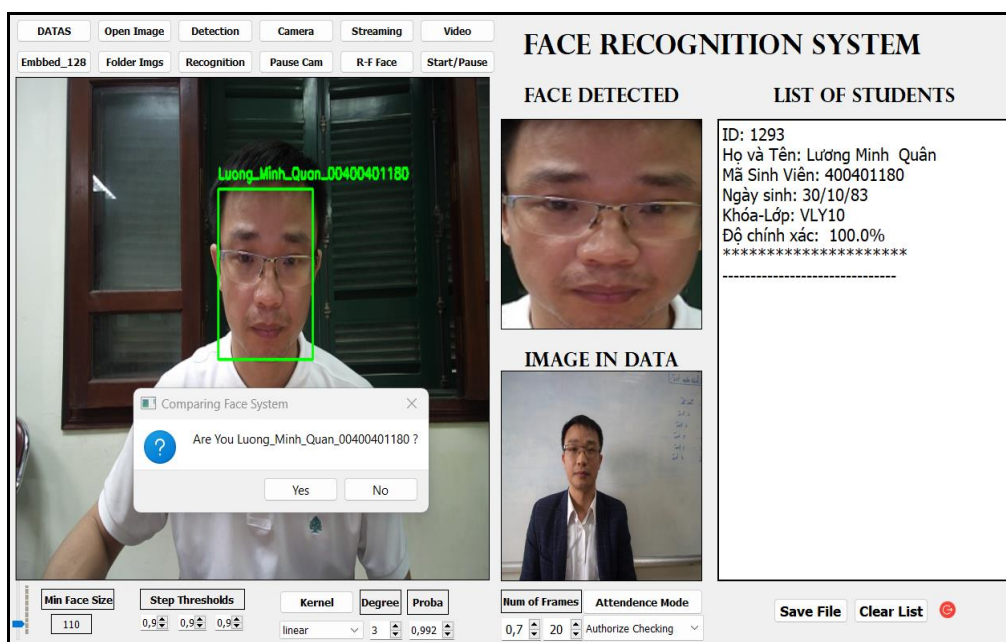
Ví dụ: Nếu $N_{frames} = 20$ và $C_{coeff} = 0,7$ thì số kết quả trả về phải cùng một nhân dạng khi phân tích từng khung hình là $N_{frames} \times C_{coeff} = 14$. Chỉ khi thỏa mãn điều kiện trên thì thông tin của người cần nhận dạng mới được hệ thống xác nhận và lưu vào bảng dữ liệu điểm danh.

Lưu ý rằng, các giá trị của N_{frames} và C_{coeff} có thể được điều chỉnh bởi người quản lý sao cho phù hợp với điều kiện thực tế khi triển khai quá trình điểm danh. Hình thức điểm danh tự động không can thiệp phù hợp để áp dụng trong các trường hợp như:

+ Giám sát thông tin tham dự lớp học của sinh viên như: bỏ giờ, đến lớp muộn; thời gian vào lớp muộn bao lâu. Từ những dữ liệu trên, đơn vị quản lý sẽ kịp thời đưa ra những cảnh báo, liên hệ với gia đình của những sinh viên

hay bỏ giờ, trốn học,... để hỗ trợ công tác quản lý sinh viên cũng như chất lượng đào tạo được nâng cao.

+ Phát hiện tình trạng gian lận trong học và thi hết học phần: mỗi khuôn mặt của sinh viên tham dự buổi học và buổi thi đều được lưu trữ và so khớp với gương mặt trong hệ thống cơ sở dữ liệu nên việc trích xuất thông tin có thể phát hiện được các sinh viên cố tình gian lận trong học và thi một cách nhanh chóng và dễ dàng.



Hình 13. Chế độ điểm danh có kiểm soát bởi giáo viên

List_Students	Number_identity	Attendance ⁺ 1	Date_of_birth	Class_hvn	Date	Time	Probability	Image_stored	Image_Composed
1 Lương Minh Quân	400401180	X	30/10/83	VLY10	04/19/23	08:09	100.0	BLOB	BLOB
2 Đoàn Duy Cẩm	6650967								
3 Lương Minh Quân	400401180								
4 Ngô Thị Thúy	6652364								
5 Nguyễn Đình Kiên	6667009								
6 Nguyễn Ngọc Long	6665591								
7 Nguyễn Ngọc Tú	6653105								
8 Nguyễn Phương Nam	6666014								
9 Nguyễn Quang Lộc	6653213								
10 Nguyễn Thế Long	6653379								
11 Nguyễn Tuấn Anh	6654317								
12 Nguyễn Vĩnh Hưng	6662500								
13 Phạm Quang Anh	6654786								
14 Trần Đình Hưởng	6667095								
15 Trần Hoài Nam	6655438								
16 Trương Võ Ngọc Văn	6660519								

Hình 14. Lưu trữ dữ liệu điểm danh của lớp học

4. KẾT LUẬN

Nhóm nghiên cứu đã phát triển thành công phần mềm nhận dạng khuôn mặt người dựa trên mạng học sâu đa nhiệm MTCNN, số hóa bởi công nghệ FaceNet và nhận diện khuôn mặt bằng thuật toán phân lớp SVM dựa trên khoảng cách giữa các véctơ số hóa trong không gian 128 chiều. Nhóm đã áp dụng phần mềm trên các nhóm sinh viên khoa CNTT tham gia thực tập Vật lý tại phòng thí nghiệm với các hình thức điểm danh khác nhau: điểm danh qua ảnh chụp; điểm danh có kiểm soát của giáo viên và điểm danh tự động. Số liệu điểm danh theo từng buổi được lưu trữ trong cơ sở dữ liệu giúp giáo viên dễ dàng thống kê thông tin tham dự lớp học của từng sinh viên.

Công nghệ điểm danh qua nhận dạng khuôn mặt đã hỗ trợ giáo viên trong việc quản lý sinh viên một cách hiệu quả, qua đó kịp thời ngăn chặn được các hình thức gian lận trong học tập như: sinh viên bỏ học và thuê người khác đi học, điểm danh; thuê người đi thi;...

Tuy nhiên, công nghệ điểm danh hiện tại chưa có khả năng phân biệt hình ảnh của người đứng trước camera và ảnh chụp của người đó được lưu lại trên các thiết bị thông minh như Ipad, Iphone,... khi mà hình ảnh này được đặt trước camera để tiến hành điểm danh. Do đó, các nghiên cứu chuyên sâu hơn khi phân tích ảnh nhiệt thu được từ camera hồng ngoại và camera thường cần tiếp tục được thực hiện trong giai đoạn tiếp theo để hoàn thiện hơn công nghệ hiện đại này.

LỜI CẢM ƠN

Để hoàn thành nghiên cứu công nghệ nhận dạng và triển khai ứng dụng công nghệ này, nhóm nghiên cứu chân thành cảm ơn sự hỗ trợ về tài chính của Học viện Nông nghiệp Việt Nam; Ban Chủ nhiệm Khoa Công nghệ thông tin; Bộ môn Vật lý và đặc biệt là các lớp sinh viên K67 Khoa Cơ - Điện, Khoa Công nghệ thông tin, K66MMT đã tham gia lấy mẫu ảnh và kiểm thử công nghệ nhận dạng trong khoảng thời gian từ tháng 11/2022 đến tháng 6/2023.

TÀI LIỆU THAM KHẢO

- Đào Quang Toàn (2017). Nhận dạng khuôn mặt trong hỗ trợ công tác quản lý tiếp dân. Luận văn Thạc sĩ Khoa học Máy tính. Đại học Thái Nguyên.
- Erik L.M., Gary B.H., Aruni R.C., Li H. & Gang H. (2016). Labeled Faces in the Wild: A Survey. In *Advances in Face Detection and Facial Image Analysis*. Springer. pp. 189-248.
- Fabian P., Gael V., Alexandre G., Vincent M., Bertrand T., Olivier G., Mathieu B., Andreas M., Joel N., Gilles L., Peter P., Ron W., Vincent D., Jake V., Alexandre P., David C., Matthieu B., Matthieu P. & Édouard D. (2011). Scikit-learn: Machine Learning in Python. *JMLR* 12, pages: 2825-2830.
- Jain V. & Erik L.M. (2010). FDDB: A Benchmark for Face Detection in Unconstrained Settings. Technical Report UM-CS-2010-009. Dept. of Computer Science, University of Massachusetts, Amherst.
- Martin K., Paul W., Peter M. R. & Horst B. (2011). Annotated Facial Landmarks in the Wild: A Large-scale, Real-world Database for Facial Landmark Localization. In *Proc. First IEEE International Workshop on Benchmarking Facial Image Analysis Technologies*.
- Ngo Tung Son, Bui Ngoc Anh, Tran Quy Ban, Le Phuong Chi, Bui Dinh Chien, Duong Xuan Hoa, Le Van Thanh, Tran Quang Huy, Le Dinh Duy & Muhammad Hassan Raza Khan (2020). Implementing CCTV-Based Attendance Taking Support System Using Deep Face Recognition: A Case Study at FPT Polytechnic College. *Symmetry*. 12(2): 307. DOI: 10.3390/sym12020307.
- Schroff. F., Kalenichenko D. & Philbin J. (2015). FaceNet: A Unified Embedding for Face Recognition and Clustering. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 815-823.
- Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V. & Rabinovich A. (2015). Going deeper with convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 1-9.
- Trần Như Ý & Nguyễn Phương Hạc (2021). Ứng dụng Haar-Cascade và MTCNN vào thuật toán ARCFACE cho hệ thống nhận dạng mặt người. *Tạp chí Công thương*. (2): 224-230.
- Yugashini I., Vidhyasri S. & Gayathri D.K. (2013). Design and Implementation of Automated Door Accessing System with Face Recognition. *International Journal of Science and Modern Engineering*. ISSN: 2319-6386. 1(12).
- Zhang Z., Zhang K., Li Z. & Qiao Y. (2016). Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks. *IEEE Signal Processing Letters*. 23(10): 1499-1503.