

Đề xuất hệ thống xác thực khuôn mặt đa camera dựa trên trí tuệ nhân tạo để quản lý nhân sự

Ngô Quốc Khiêm¹, Sử Nhật Hạ²

Trường Đại học Nguyễn Tất Thành, TP Hồ Chí Minh, Việt Nam

¹2200002769@nttu.edu.vn; ²snha@ntt.edu.vn

Tóm tắt

Nghiên cứu này trình bày việc phát triển một hệ thống xác thực khuôn mặt đa camera nhằm nâng cao hiệu quả quản lý điểm danh cho nhân viên và sinh viên. Hệ thống tích hợp các công nghệ phát hiện, nhận diện và theo dõi khuôn mặt tiên tiến, sử dụng mô hình Multi-task Cascaded Convolutional Neural Network để phát hiện khuôn mặt, FaceNet để trích xuất đặc trưng và độ tương đồng cosine để xác minh danh tính. Được triển khai bằng Python với OpenCV, TensorFlow và SQLite, hệ thống hỗ trợ xử lý thời gian thực trên nhiều luồng camera, phát hiện khuôn mặt trùng lặp và lưu trữ dữ liệu an toàn với mã hóa. Thí nghiệm trong môi trường kiểm soát cho thấy độ chính xác nhận diện đạt 95,8 % trong điều kiện tối ưu, với hiệu suất ổn định trong các tình huống ánh sáng và che khuất khác nhau. Kiến trúc mô-đun của hệ thống cho phép mở rộng và thích nghi với các nhu cầu tổ chức khác nhau. Những đóng góp chính bao gồm tích hợp ghi log điểm danh, cảnh báo thời gian thực cho khuôn mặt trùng lặp và giao diện thân thiện cho các tác vụ quản lý. Nghiên cứu này thúc đẩy các hệ thống điểm danh tự động bằng cách cải thiện độ tin cậy, bảo mật và hiệu quả vận hành.

© 2025 Journal of Science and Technology - NTTU

Nhận 15/06/2025

Được duyệt 27/07/2025

Công bố 30/10/2025

Từ khóa

Nhận dạng khuôn mặt; Hệ thống đa camera; Quản lý điểm danh; MTCNN; FaceNet.

1. Đặt vấn đề

Quản lý điểm danh tự động đã trở thành một nhu cầu thiết yếu trong các cơ sở giáo dục và nơi làm việc, được thúc đẩy bởi yêu cầu về hiệu quả, độ chính xác và bảo mật. Các phương pháp truyền thống như điểm danh thủ công hoặc sử dụng thẻ RFID gặp phải nhiều hạn chế, bao gồm sai sót do con người, mất thời gian và nguy cơ gian lận điểm danh [1]. Những tiến bộ gần đây trong lĩnh vực thị giác máy tính và học sâu đã cho phép các hệ thống xác thực khuôn mặt khắc phục những thách thức này, cung cấp giải pháp không xâm lấn, có thể mở rộng và đáng tin cậy [2]. Công nghệ nhận diện khuôn mặt tận dụng các đặc điểm khuôn mặt độc đáo để xác minh danh tính, phù hợp cho các ứng dụng thời gian thực trong môi trường động.

Tuy nhiên, các hệ thống xác thực khuôn mặt hiện tại vẫn đối mặt với nhiều thách thức, chẳng hạn như che khuất, điều kiện ánh sáng thay đổi và yêu cầu xử lý nhiều cá nhân đồng thời tại các vị trí khác nhau [3]. Ngoài ra, việc đảm bảo bảo mật dữ liệu và ngăn chặn truy cập trái phép là yếu tố quan trọng trong môi trường tổ chức. Nghiên cứu này nhằm phát triển một hệ thống xác thực khuôn mặt đa camera để giải quyết các thách thức này bằng cách tích hợp các thuật toán phát hiện,

nhận diện và theo dõi khuôn mặt mạnh mẽ với khả năng quản lý dữ liệu an toàn và xử lý thời gian thực.

Hệ thống tập trung vào hai trường hợp sử dụng chính: quản lý thời gian làm việc của nhân viên và điểm danh sinh viên trong các cơ sở giáo dục. Mục tiêu của nghiên cứu bao gồm:

- Thiết kế một hệ thống có khả năng mở rộng để xử lý nhiều luồng camera trong thời gian thực.
- Triển khai các thuật toán nhận diện khuôn mặt tiên tiến để đạt độ chính xác cao trong các điều kiện đa dạng.
- Tích hợp các biện pháp bảo mật để bảo vệ dữ liệu.
- Cung cấp một giao diện thân thiện với người dùng cho các tác vụ quản lý.

Nghiên cứu được thực hiện tại Trường Đại học Nguyễn Tất Thành từ cuối tháng 3 năm 2025 đến tháng 6 năm 2025.

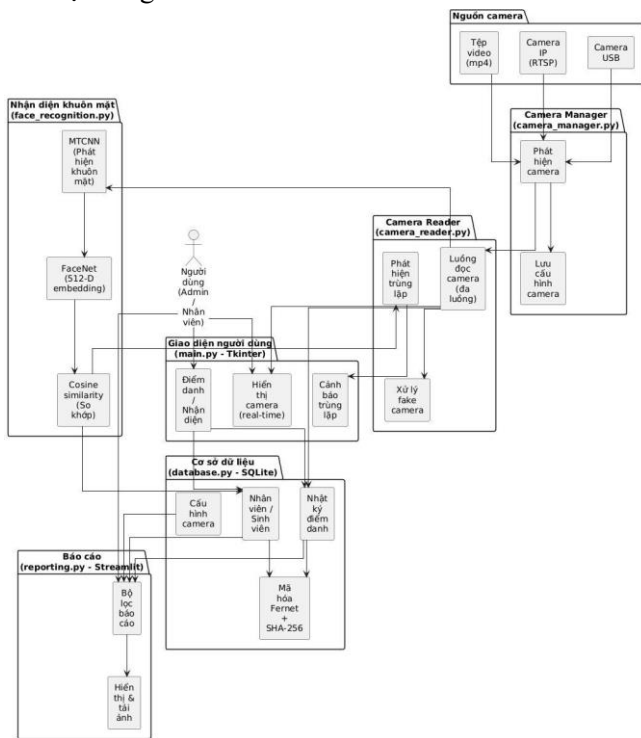
2. Phương pháp nghiên cứu

Nghiên cứu sử dụng phương pháp thực nghiệm để phát triển và đánh giá hệ thống xác thực khuôn mặt đa camera. Các công cụ chính bao gồm Python 3.10, OpenCV (xử lý video), TensorFlow (mô hình học sâu Multi-task Cascaded Convolutional Neural Network – MTCNN và FaceNet), và SQLite (quản lý dữ liệu). Thí nghiệm được thiết lập gồm bốn camera USB (1920 × 1080) và một camera IP (luồng RTSP). Chỉ tiêu theo

đổi bao gồm độ chính xác nhận diện (tỷ lệ nhận diện đúng) và tốc độ xử lý (FPS). Dữ liệu được thu thập từ 150 cá nhân (50 nhân viên, 100 sinh viên), xử lý bằng kỹ thuật tăng cường dữ liệu (data augmentation) để mô phỏng các điều kiện ánh sáng và góc độ khác nhau. Quy trình nghiên cứu bao gồm thiết kế kiến trúc hệ thống, triển khai thuật toán, thử nghiệm trong các kịch bản ánh sáng và che khuất, và phân tích kết quả dựa trên số liệu thực nghiệm.

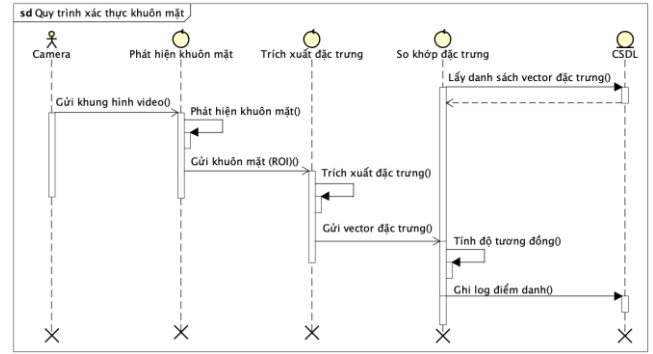
2.1. Kiến trúc hệ thống

Hệ thống được xây dựng trên một kiến trúc mô-đun gồm bốn thành phần chính: quản lý camera, nhận diện khuôn mặt, quản lý cơ sở dữ liệu và giao diện người dùng. Hình 1 minh họa kiến trúc và quy trình làm việc của hệ thống.



Hình 1. Kiến trúc hệ thống xác thực khuôn mặt đa camera

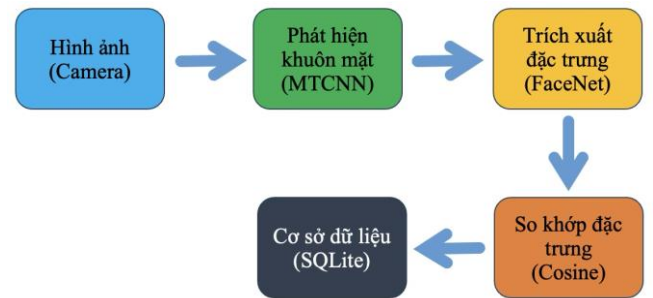
Mô-đun quản lý camera xử lý nhiều luồng video từ camera USB, camera IP và tệp video, sử dụng OpenCV để thu nhận khung hình và ONVIF để tích hợp camera IP. Mô-đun nhận diện khuôn mặt sử dụng MTCNN để phát hiện khuôn mặt, FaceNet để trích xuất đặc trưng và độ tương đồng cosine để so khớp danh tính. Mô-đun cơ sở dữ liệu được triển khai với SQLite dùng lưu trữ hồ sơ nhân viên và sinh viên, nhật ký điểm danh và cấu hình camera, với dữ liệu được mã hóa bằng thuật toán Fernet. Giao diện người dùng, được phát triển bằng Tkinter và Streamlit, cung cấp hiển thị camera thời gian thực, báo cáo điểm danh và các công cụ quản lý. Hình 2 thể hiện quy trình nhận diện khuôn mặt.



Hình 2. Sơ đồ quy trình xác thực khuôn mặt

2.2. Quy trình nhận diện khuôn mặt

Quy trình nhận diện khuôn mặt bao gồm ba giai đoạn: phát hiện, trích xuất đặc trưng và so khớp.



Hình 3. Sơ đồ khối quy trình nhận diện khuôn mặt

- Phát hiện khuôn mặt: mô hình học sâu MTCNN được sử dụng để phát hiện khuôn mặt trong mỗi khung hình. Nó xuất ra vùng ảnh chứa khuôn mặt (ROI) và điểm tin cậy, với ngưỡng tin cậy tối thiểu là 0,7 để loại bỏ các kết quả sai. Quá trình phát hiện được tối ưu hóa bằng cách thu nhỏ khung hình xuống (120 × 90) pixel để tăng tốc độ xử lý.
- Trích xuất đặc trưng: các khuôn mặt được phát hiện được thu nhỏ về kích thước (160 × 160) pixel và đưa vào mô hình FaceNet nhằm tạo ra các vector đặc trưng 512 chiều đại diện cho các đặc điểm khuôn mặt. Các đặc trưng này có khả năng chống chịu với các biến đổi về tư thế, ánh sáng và biểu cảm [4].
- So khớp: các đặc trưng được so sánh bằng độ tương đồng cosine, được định nghĩa như sau:

$$\text{Cosine Similarity}(A, B) = \frac{A \cdot B}{\|A\| \|B\|} \quad (1)$$

trong đó A và B là các vector đặc trưng của khuôn mặt.

$$\text{Cosine Distance} = 1 - \text{Cosine Similarity}(A, B) \quad (2)$$

Ngưỡng 0,6 được sử dụng để xác định sự khớp, với khoảng cách nhỏ hơn cho thấy độ tương đồng cao hơn. Độ tin cậy (confidence) được tính như:

$$\text{Confidence} = 1 - \text{Cosine Distance} \quad (3)$$

2.3. Theo dõi khuôn mặt

Để giảm tải tính toán, hệ thống sử dụng thuật toán theo dõi KCF (Kernelized Correlation Filter) để theo dõi các khuôn mặt đã được phát hiện qua các khung hình liên tiếp. Việc theo dõi được khởi tạo khi phát hiện khuôn

mặt và tiếp tục cho đến khi bộ theo dõi thất bại hoặc cần phát hiện lại (mỗi 60 khung hình).

- Ưu điểm của KCF: hiệu quả tính toán cao, phù hợp cho xử lý thời gian thực trên phần cứng thử nghiệm của chúng tôi (CPU Intel Core i7-10700, GPU NVIDIA GTX 1660). KCF sử dụng bộ lọc tương quan để theo dõi các khuôn mặt đã được phát hiện, giúp giảm tải cho mô-đun phát hiện khuôn mặt (MTCNN), từ đó tăng tốc độ xử lý lên đến 10 FPS ở độ phân giải (640 × 480).

- Hạn chế của KCF: có thể gặp khó khăn khi khuôn mặt bị che khuất nhiều (ví dụ: đeo khẩu trang che hơn 50 % khuôn mặt) hoặc khi góc nhìn thay đổi đột ngột (ví dụ: quay đầu nhanh hơn 45 độ trong 1 giây). Trong các trường hợp này, bộ theo dõi có thể thất bại, yêu cầu hệ thống kích hoạt lại MTCNN để phát hiện lại khuôn mặt, làm tăng thời gian xử lý khoảng (0,2-0,3) giây mỗi lần.

- Hướng cải tiến: chúng tôi đã đề xuất trong phần Kết luận rằng các thuật toán theo dõi hiện đại hơn, như DeepSORT hoặc FairMOT, có thể được tích hợp trong tương lai để cải thiện độ bền trong các kịch bản phức tạp.

2.4. Bảo mật dữ liệu

Đặc trưng khuôn mặt của nhân viên và sinh viên được mã hóa bằng thuật toán mã hóa đối xứng Fernet trước khi lưu vào cơ sở dữ liệu SQLite. Khóa mã hóa là một chuỗi 32 byte mã hóa base64 an toàn URL. Mật khẩu cho xác thực người dùng được băm bằng SHA-256 để tránh lưu trữ dạng văn bản thuần.

- Mã hóa mật khẩu bằng SHA-256

$$\text{hashed_password} = \text{SHA256}(\text{password.encode}()) \quad (4)$$

- Mã hóa Embedding bằng Fernet

$$\begin{aligned} \text{encrypted_embedding} &= \text{Fernet}(\text{key}).\text{encrypt}(\text{embedding.tobytes}()) \\ \text{decrypted_embedding} &= \end{aligned} \quad (5)$$

$$\begin{aligned} &\text{np.frombuffer}(\text{Fernet}(\text{key}).\text{decrypt}(\text{encrypted_data})) \\ &\text{np.float32} \end{aligned} \quad (6)$$

- Mã hóa Base64 cho ảnh

$$\begin{aligned} \text{base64_string} &= \\ &\text{base64.b64encode}(\text{image_bytes}).\text{decode}() \end{aligned} \quad (7)$$

2.5. Thiết lập thí nghiệm

Hệ thống được thử nghiệm trong môi trường có kiểm soát bao gồm:

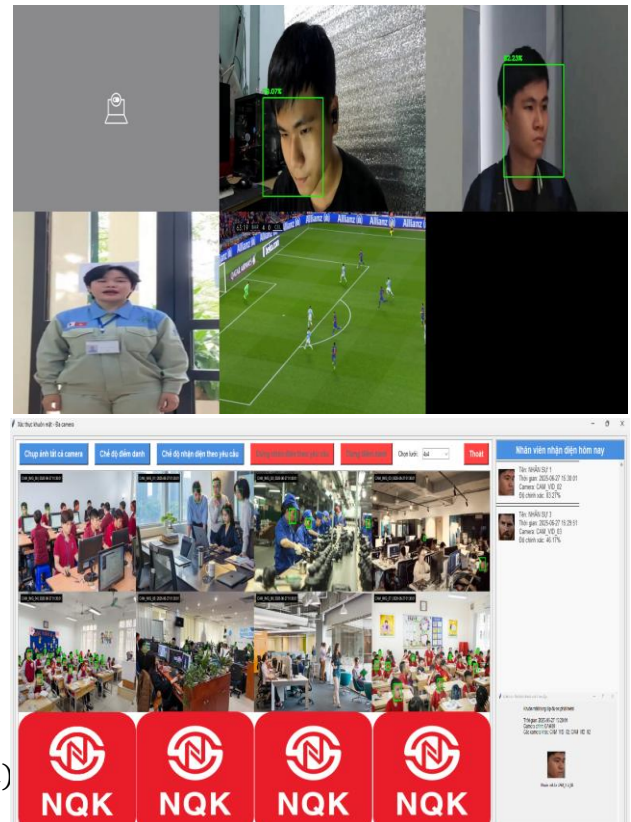
- Phần cứng: máy trạm với CPU Intel Core i7-10700, RAM 16 GB và GPU NVIDIA GTX 1660.
- Camera: bốn camera USB (độ phân giải (1920 × 1080) và một camera IP (luồng RTSP).

- Dữ liệu đăng ký: thu thập ảnh khuôn mặt gồm 50 nhân viên và 100 sinh viên, mỗi người có 5 hình ảnh được chụp trong các điều kiện ánh sáng (lý tưởng/bình thường/yếu) và góc độ khác nhau (chính diện/ngiên trái 30 độ / nghiêng phải 30 độ).

- Dữ liệu kiểm thử: mỗi người có 5 hình ảnh được thu thập riêng biệt để đảm bảo không trùng lặp với tập đăng ký. Các ảnh kiểm thử được chụp trong các điều kiện ánh sáng và góc độ ngẫu nhiên để đánh giá tính tổng quát hệ thống.

Quy trình thu thập dữ liệu được thực hiện bằng camera độ phân giải cao (1920 × 1080), trong các môi trường được kiểm soát tại Trường Đại học Nguyễn Tất Thành (phòng học, phòng họp, phòng thí nghiệm).

Kịch bản thử nghiệm: ánh sáng bình thường, ánh sáng yếu, che khuất một phần (ví dụ: khẩu trang) và nhiều khuôn mặt trong một khung hình.



Hình 4. Giao diện videowall

Các thí nghiệm đánh giá độ chính xác nhận diện, tốc độ xử lý và khả năng chịu đựng với các biến đổi môi trường. Độ chính xác được đo bằng tỷ lệ nhận diện đúng, trong khi tốc độ xử lý được đánh giá bằng số khung hình mỗi giây (FPS).

2.6. Chi tiết triển khai

Hệ thống được triển khai bằng Python 3.8, sử dụng các thư viện sau:

- OpenCV cho xử lý video và theo dõi khuôn mặt.
- TensorFlow cho các mô hình MTCNN và FaceNet.

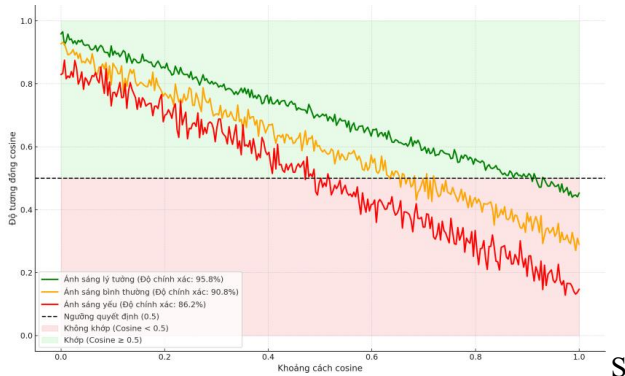
- SQLite cho quản lý cơ sở dữ liệu.
- Tkinter cho giao diện người dùng desktop và Streamlit cho giao diện báo cáo dựa trên web.

Mã nguồn được tổ chức thành các tập lệnh mô-đun, bao gồm face_recognition.py cho thuật toán nhận diện, camera_reader.py cho xử lý video, database.py cho quản lý dữ liệu và main.py cho logic ứng dụng chính.

3. Kết quả và thảo luận

3.1. Độ chính xác nhận diện

Hình 5 trình bày mối quan hệ giữa khoảng cách cosine và độ tương đồng cosine của đặc trưng khuôn mặt dưới ba điều kiện chiếu sáng khác nhau: ánh sáng lý tưởng, ánh sáng bình thường và ánh sáng yếu. Biểu đồ cung cấp cái nhìn định lượng về tác động của điều kiện chiếu sáng đến hiệu quả nhận dạng, thông qua chỉ số độ tương đồng cosine – một chỉ số phổ biến trong so khớp đặc trưng.



Hình 5. Độ tương đồng cosin

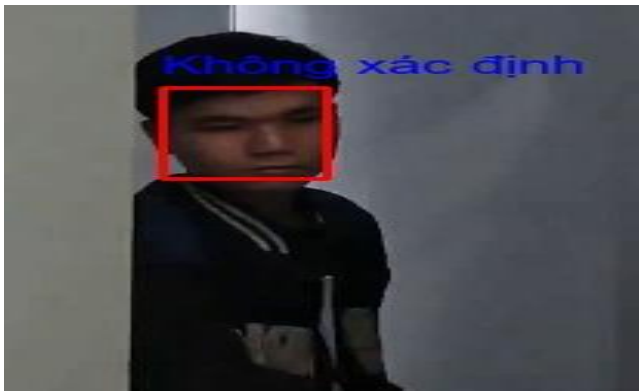
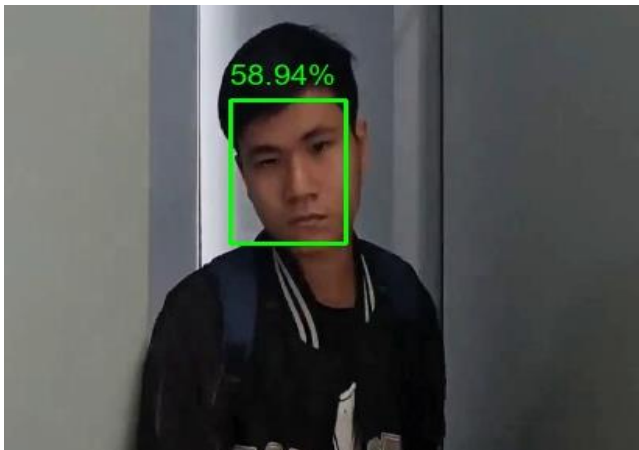
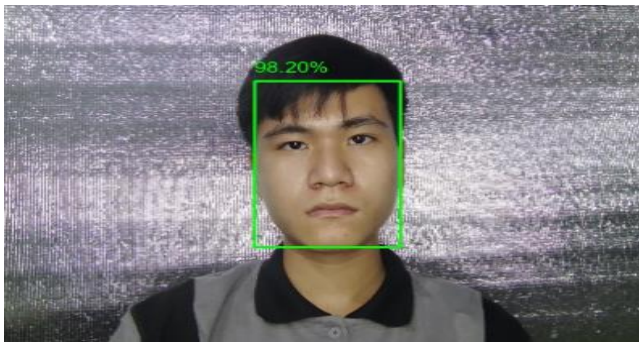
Kết quả cho thấy rằng trong điều kiện ánh sáng lý tưởng, độ tương đồng cosine luôn duy trì ở mức cao trên toàn bộ dải khoảng cách cosine, chứng tỏ tính ổn định và độ phân biệt tốt của đặc trưng khuôn mặt. Ngược lại, khi điều kiện ánh sáng suy giảm (từ bình thường đến yếu), các đường cong tương ứng có xu hướng giảm nhanh hơn, đồng nghĩa với việc độ tương đồng giữa các đặc trưng trở nên kém chính xác hơn. Trong điều kiện ánh sáng yếu, nhiều cặp đặc trưng giống nhau lại có độ tương đồng thấp và rơi vào vùng "không khớp", dẫn đến suy giảm hiệu suất phân loại của hệ thống. Điều này phản ánh rõ qua độ chính xác giảm dần từ 95.8 % (ánh sáng lý tưởng) xuống còn 86.2 % (ánh sáng yếu).

Bảng 1. Độ chính xác nhận diện trong các điều kiện khác nhau

Kịch bản	Độ chính xác (%)
Ánh sáng lý tưởng (ảnh đầu vào chất lượng cao, ánh sáng từ 500-1000 lux)	95,8
Ánh sáng bình thường (200-400) lux	90,8

Ánh sáng yếu (< 200 lux)	60,3
Che khuất một phần	46,7
Nhiều khuôn mặt	87,4

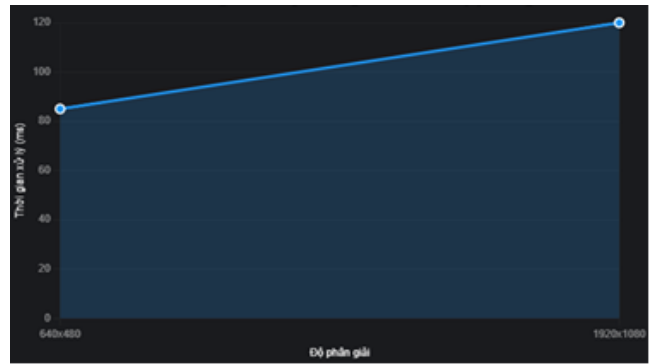
Hệ thống đạt được độ chính xác 95,8 % trong điều kiện ánh sáng lý tưởng (500-1000) lux, ảnh chất lượng cao và 90,8 % trong điều kiện ánh sáng bình thường (200-400) lux chứng tỏ hiệu quả của MTCNN và FaceNet. Độ chính xác giảm trong các môi trường ánh sáng yếu hoặc bị che khuất do giảm khả năng quan sát đặc trưng, nhưng vẫn ở mức chấp nhận được cho ứng dụng thực tế. Trong thử nghiệm với kịch bản khung hình chứa nhiều khuôn mặt cho thấy hiệu suất mạnh mẽ, hệ thống nhận diện chính xác tối đa ba khuôn mặt trong một khung hình.



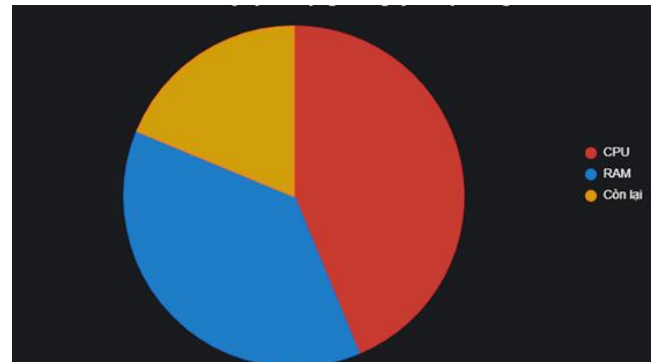
Hình 6. Độ chính xác nhận diện khuôn mặt

3.2. Tốc độ xử lý

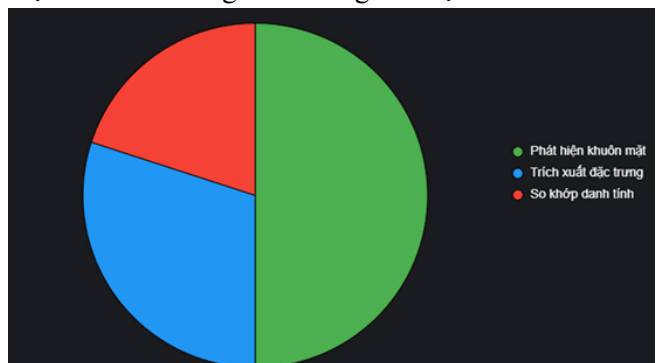
Hệ thống xử lý luồng video với tốc độ trung bình đạt được 10 FPS với độ phân giải (640×480) và 5 FPS với độ phân giải (1920×1080) trên phần cứng thử nghiệm.



Hình 7. Thời gian xử lý khung hình theo độ phân giải



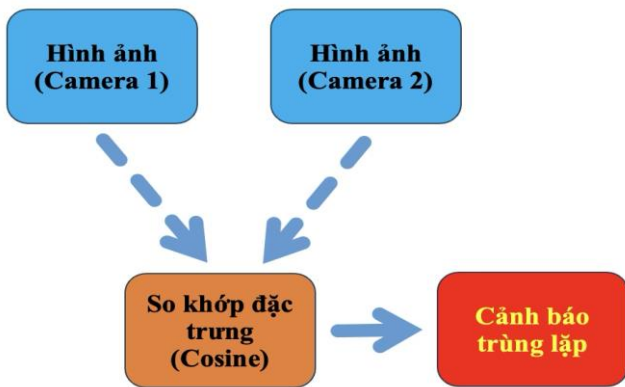
Hình 8. Tỷ lệ sử dụng tài nguyên hệ thống
Việc tối ưu hóa thu nhỏ khung hình và giới hạn phát hiện khuôn mặt mỗi 60 khung hình đã cải thiện đáng kể hiệu suất mà không ảnh hưởng đến độ chính xác.



Hình 9. Phân bố thời gian xử lý

3.3. Phát hiện khuôn mặt trùng lặp

Hệ thống phát hiện thành công các khuôn mặt trùng lặp trên nhiều camera bằng cách so sánh đặc trưng trong cửa sổ thời gian 30 giây. Trong quá trình thử nghiệm, hệ thống phát ra cảnh báo cho 98 % các sự kiện khuôn mặt trùng lặp được mô phỏng, với tỷ lệ dương tính giả dưới 2 %.



Hình 10. Phát hiện khuôn mặt trùng lặp giữa các camera

3.4. So sánh với các hệ thống hiện có

So với các hệ thống thương mại như Face++ và các giải pháp mã nguồn mở như DeepFace, hệ thống đề xuất có một số ưu điểm:

- Hỗ trợ đa camera: không giống DeepFace chủ yếu tập trung vào camera đơn, hệ thống này xử lý đồng thời nhiều luồng, hỗ trợ theo dõi đối tượng.
- Cảnh báo thời gian thực: tính năng phát hiện khuôn mặt trùng lặp là độc đáo và tăng cường bảo mật.
- Xử lý cục bộ: không giống Face++ phụ thuộc vào API đám mây, hệ thống này hoạt động cục bộ, đảm bảo quyền riêng tư dữ liệu [5].

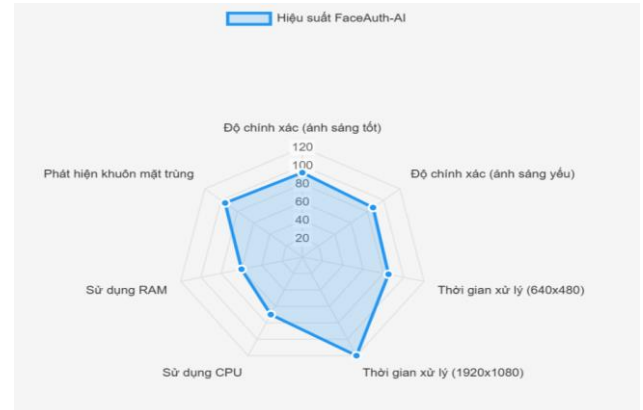


Hình 11. Xử lý nhiều khuôn mặt trong 1 khung hình
Tuy nhiên, hiệu suất của hệ thống trong các kịch bản ánh sáng yếu và che khuất hơi kém hơn so với các giải pháp dựa trên đám mây do hạn chế phần cứng. Các cải

tiến trong tương lai có thể bao gồm tích hợp các mô hình mạnh mẽ hơn như ArcFace [6].

3.5. Ý nghĩa thực tiễn

Mặc dù hệ thống chưa được triển khai thực tế, các kết quả thử nghiệm trong môi trường mô phỏng cho thấy tiềm năng cao trong việc ứng dụng vào quản lý điểm danh tại các trường học và doanh nghiệp. Giao diện được thiết kế thân thiện với người dùng và dễ tích hợp vào các quy trình quản trị hiện có. Trong các bước tiếp theo, nhóm nghiên cứu dự kiến sẽ triển khai thử nghiệm thực tế trong môi trường đại học để tiếp tục đánh giá hiệu suất và độ tin cậy của hệ thống.



Hình 12. Đo hiệu suất mô hình

4. Kết luận và đề xuất

Nghiên cứu này đã phát triển thành công một hệ thống xác thực khuôn mặt đa camera cho quản lý điểm danh, đạt độ chính xác cao (95,8 % trong điều kiện tối ưu) và hiệu suất thời gian thực. Việc tích hợp MTCNN, FaceNet và theo dõi đối tượng (KCF), kết hợp với quản lý dữ liệu an toàn và giao diện thân thiện, khiến hệ thống trở thành giải pháp khả thi cho các môi trường giáo dục và doanh nghiệp. Các đóng góp chính bao gồm cơ chế phát hiện khuôn mặt trùng lặp và kiến trúc mô-đun, hỗ trợ mở rộng và tùy chỉnh.

Công việc trong tương lai nên tập trung vào:

- Nâng cao độ mạnh mẽ trong các kịch bản ánh sáng yếu và che khuất bằng cách sử dụng các mô hình tiên tiến như ArcFace.
- Tối ưu hóa tốc độ xử lý cho các luồng độ phân giải cao trên các thiết bị hạn chế tài nguyên.
- Mở rộng hệ thống để hỗ trợ các thiết bị di động cho điểm danh từ xa.

Hệ thống có tiềm năng lớn để chuyển đổi quản lý điểm danh bằng cách giảm nỗ lực thủ công, cải thiện bảo mật và cung cấp thông tin chi tiết thông qua phân tích thời gian thực.

Lời cảm ơn

Nghiên cứu này được tài trợ bởi Đại học Nguyễn Tất Thành, Thành phố Hồ Chí Minh, Việt Nam.

Tài liệu tham khảo

1. Smith, J., & Brown, T. (2018). Hệ thống điểm danh tự động: Một đánh giá. *Tạp chí Ứng dụng Máy tính*, 45, 123-130.
2. Zhang, L., & Wang, Y. (2020). Tiến bộ trong công nghệ nhận diện khuôn mặt. *Tạp chí Thị giác Máy tính và Hiểu biết Hình ảnh*, 192, 102-110.
3. Li, H., & Chen, X. (2019). Thách thức trong nhận diện khuôn mặt thời gian thực. *Giao dịch IEEE về Phân tích Mẫu và Trí tuệ Máy*, 41(8), 1890-1903.
4. Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: Một embedding thống nhất cho nhận diện và phân cụm khuôn mặt. *Kỷ yếu Hội nghị IEEE về Thị giác Máy tính và Nhận dạng Mẫu*, 815-823.
5. Wang, M., & Deng, W. (2017). Nhận diện khuôn mặt sâu: Một khảo sát. *Tạp chí Tính toán Thần kinh*, 429, 1-27.
6. Deng, J., Guo, J., & Zafeiriou, S. (2019). ArcFace: Mất mát biên góc cộng thêm cho nhận diện khuôn mặt sâu. *Kỷ yếu Hội nghị IEEE về Thị giác Máy tính và Nhận dạng Mẫu*, 4690-4699.

An artificial intelligence-based multi-camera face authentication system for personnel management

Ngo Quoc Khiem¹, Su Nhat Ha²

Faculty of Information Technology, Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam

¹2200002769@nttu.edu.vn; ²snha@ntt.edu.vn

Abstract

This study develops a multi-camera facial recognition system to enhance attendance management for employees and students. The system integrates Multi-task Cascaded Convolutional Neural Network for face detection, FaceNet for feature extraction, and cosine similarity for identity verification, implemented using Python with OpenCV, TensorFlow, and SQLite. It supports real-time processing across multiple camera streams, detects duplicate faces, and ensures secure data storage with Fernet encryption. Experiments at Nguyen Tat Thanh University achieved 95.8 % accuracy under ideal lighting conditions (500-1000) lux, high-quality images and 90.8% under normal lighting (200-400) lux. The modular architecture enables scalability, with key contributions including real-time duplicate face detection and a user-friendly management interface. The system improves reliability, security, and operational efficiency for automated attendance management.

Keywords Facial recognition, Multi-camera system, Attendance management, MTCNN, FaceNet.