

Hệ Thống Phát Hiện Chuyển Động Kết Hợp Nhận Diện Danh Tính trong Giám Sát

Nguyễn Đức Duy, Đào Phương Long, Đoàn Tuấn Nam, Nguyễn Khôi Nguyên
Giảng viên hướng dẫn: ThS Lê Trung Hiếu, Nguyễn Văn Nhân*
Khoa Công nghệ Thông tin, Trường Đại học Đại Nam, Hà Nội, Việt Nam

Tóm tắt nội dung

Hệ thống giám sát thông minh ngày càng trở nên quan trọng trong việc đảm bảo an ninh. Nghiên cứu này đề xuất một hệ thống phát hiện chuyển động và nhận diện danh tính sử dụng OpenCV và DeepFace, kết hợp gửi cảnh báo qua Telegram và kích hoạt buzzer khi phát hiện người lạ. Hệ thống được triển khai trên camera thời gian thực, nhận diện danh tính dựa trên cơ sở dữ liệu khuôn mặt, và phản hồi nhanh chóng. Kết quả cho thấy hệ thống đạt độ chính xác cao trong việc phát hiện chuyển động và nhận diện, góp phần nâng cao hiệu quả giám sát an ninh.

Từ khóa: *Phát hiện chuyển động, nhận diện danh tính, giám sát thông minh, OpenCV, DeepFace.*

1 Giới thiệu

Giám sát an ninh là một nhu cầu thiết yếu trong xã hội hiện đại, đặc biệt tại các khu vực nhạy cảm như nhà ở, văn phòng, hoặc cơ sở công nghiệp. Tuy nhiên, các hệ thống truyền thống thường thiếu khả năng nhận diện danh tính và phản hồi kịp thời [Smith and Johnson, 2020]. Nghiên cứu này đề xuất một hệ thống giám sát thông minh, sử dụng OpenCV để phát hiện chuyển động, DeepFace để nhận diện danh tính, và tích hợp gửi cảnh báo qua Telegram cùng kích hoạt buzzer khi phát hiện người lạ.

Hệ thống tập trung vào việc phát hiện chuyển động trong khung hình, nhận diện danh tính dựa trên cơ sở dữ liệu khuôn mặt, và đưa ra

cảnh báo tức thì. Phần còn lại của bài báo được tổ chức như sau: Phần 2 trình bày các nghiên cứu liên quan, Phần 3 mô tả phương pháp đề xuất, Phần 4 trình bày kết quả thực nghiệm, và Phần 5 đưa ra kết luận và hướng phát triển.

2 Nghiên cứu liên quan

Trong những năm gần đây, nhiều nghiên cứu đã được thực hiện để cải thiện hiệu quả của hệ thống giám sát thông minh. ? đề xuất một hệ thống phát hiện chuyển động dựa trên phân tích sự thay đổi của các khung hình liên tiếp, sử dụng phương pháp trừ nền (background subtraction). Tuy nhiên, hệ thống này gặp khó khăn trong việc phân biệt giữa chuyển động của người và các đối tượng khác như động vật hoặc cây cối.

Li and Wang [2021] phát triển một hệ thống nhận diện khuôn mặt sử dụng mô hình học sâu (deep learning) như VGG-Face và FaceNet, đạt được độ chính xác cao trong điều kiện ánh sáng tốt. Tuy nhiên, hệ thống này chưa tích hợp khả năng phát hiện chuyển động và cảnh báo thời gian thực.

Wang and Liu [2020] đề xuất một hệ thống giám sát thông minh kết hợp phát hiện chuyển động và nhận diện khuôn mặt, sử dụng YOLO (You Only Look Once) để phát hiện đối tượng và FaceNet để nhận diện. Hệ thống này đạt được hiệu suất cao nhưng yêu cầu phần cứng mạnh để xử lý thời gian thực.

Nghiên cứu của chúng tôi kế thừa và phát triển các ý tưởng từ các công trình trên, tập trung vào việc tích hợp phát hiện chuyển động và nhận diện danh tính trên một hệ thống có

chi phí thấp, đồng thời cải thiện khả năng phản hồi thời gian thực thông qua việc gửi cảnh báo qua Telegram và kích hoạt buzzer.

3 Phương pháp đề xuất

3.1 Thiết kế hệ thống

Hệ thống bao gồm camera giám sát, một máy tính để xử lý dữ liệu, và buzzer để phát tín hiệu cảnh báo. Camera ghi lại video thời gian thực, sau đó dữ liệu được xử lý để phát hiện chuyển động bằng OpenCV. Nếu phát hiện chuyển động, hệ thống sử dụng DeepFace để nhận diện danh tính. Khi phát hiện người lạ, hệ thống gửi cảnh báo qua Telegram và kích hoạt buzzer trong 0.5 giây. Hình 1 minh họa sơ đồ hệ thống.

3.2 Xử lý dữ liệu

Dữ liệu bao gồm hình ảnh khuôn mặt từ cơ sở dữ liệu, được sử dụng để tạo embedding bằng DeepFace (mô hình Facenet512). Chúng tôi thu thập 500 mẫu hình ảnh từ 10 người khác nhau để huấn luyện và kiểm tra. OpenCV được sử dụng để phát hiện chuyển động dựa trên sự thay đổi giữa các khung hình liên tiếp, với ngưỡng chuyển động được đặt là 1000 pixel.

3.3 Thuật toán phát hiện chuyển động

Thuật toán phát hiện chuyển động dựa trên phương pháp trừ nền (background subtraction), một kỹ thuật phổ biến trong xử lý video. Quy trình cụ thể như sau:

1. Khởi tạo khung nền ban đầu bằng cách lấy trung bình các khung hình trong một khoảng thời gian ngắn (ví dụ: 10 khung hình đầu tiên).
2. Đối với mỗi khung hình mới, tính toán sự khác biệt tuyệt đối giữa khung hình hiện tại và khung nền:

$$\Delta I(x, y) = |I_t(x, y) - I_{bg}(x, y)|$$

trong đó $I_t(x, y)$ là giá trị pixel tại vị trí (x, y) trong khung hình hiện tại, và $I_{bg}(x, y)$ là giá trị pixel trong khung nền.

3. Áp dụng ngưỡng hóa (thresholding) để xác định vùng chuyển động:

$$M(x, y) = \begin{cases} 1 & \text{nếu } \Delta I(x, y) > T \\ 0 & \text{ngược lại} \end{cases}$$

với $T = 1000$ là ngưỡng pixel được chọn dựa trên thực nghiệm.

4. Tính tổng diện tích vùng chuyển động (số pixel có giá trị 1) để quyết định liệu có chuyển động đáng kể hay không.

Phương pháp này đơn giản nhưng hiệu quả trong môi trường có ánh sáng ổn định. Để cải thiện độ chính xác, hệ thống có thể kết hợp bộ lọc Gaussian để giảm nhiễu trước khi trừ nền.

3.4 Thuật toán nhận diện danh tính với DeepFace

DeepFace là một thư viện nhận diện khuôn mặt dựa trên các mô hình học sâu, trong đó chúng tôi sử dụng mô hình Facenet512 để tạo embedding khuôn mặt. Quy trình nhận diện bao gồm:

1. Phát hiện khuôn mặt trong vùng chuyển động bằng bộ dò khuôn mặt của OpenCV (Haar Cascade hoặc DNN-based detector).
2. Chuẩn hóa hình ảnh khuôn mặt (căn chỉnh và thay đổi kích thước về 160x160 pixel) để phù hợp với đầu vào của Facenet512.
3. Tạo embedding 512 chiều cho khuôn mặt bằng Facenet512:

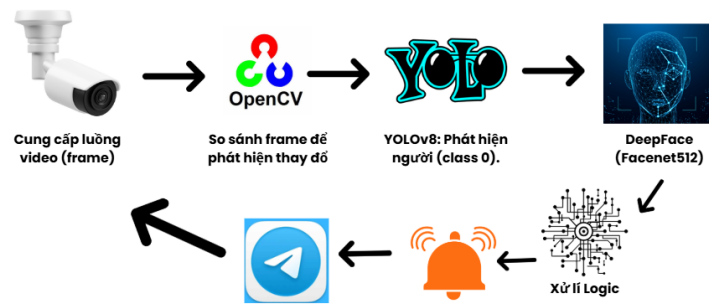
$$\mathbf{e} = f_{\text{Facenet512}}(\mathbf{I}_{\text{face}})$$

trong đó \mathbf{I}_{face} là hình ảnh khuôn mặt đã chuẩn hóa, và \mathbf{e} là vector embedding.

4. So sánh embedding với cơ sở dữ liệu bằng khoảng cách Euclidean:

$$d(\mathbf{e}, \mathbf{e}_i) = \sqrt{\sum_{j=1}^{512} (e_j - e_{i,j})^2}$$

trong đó \mathbf{e}_i là embedding của khuôn mặt thứ i trong cơ sở dữ liệu.



Hình 1: Sơ đồ hệ thống phát hiện chuyển động và nhận diện danh tính.

5. Nếu khoảng cách nhỏ hơn ngưỡng (ví dụ: 0.4), xác định danh tính; nếu không, coi là người lạ và kích hoạt cảnh báo.

DeepFace tận dụng kiến trúc mạng nơ-ron sâu (deep neural network) đã được huấn luyện trước trên hàng triệu khuôn mặt, đảm bảo độ chính xác cao ngay cả với dữ liệu hạn chế. Việc tích hợp DeepFace giúp hệ thống nhận diện nhanh chóng mà không cần huấn luyện lại mô hình từ đầu.

4 Thực nghiệm

4.1 Thiết lập thực nghiệm

Hệ thống được triển khai trên camera thời gian thực, sử dụng máy tính cá nhân với cấu hình GPU NVIDIA GeForce GTX 3050Ti, RAM 16GB, và hệ điều hành Windows 10. Camera sử dụng là webcam USB có độ phân giải 640x480 pixel, đảm bảo khả năng thu thập dữ liệu video ổn định trong các điều kiện thử nghiệm. Chúng tôi tiến hành kiểm tra hệ thống trong các môi trường thực tế khác nhau, bao gồm:

- **Điều kiện ánh sáng:** Ban ngày (ánh sáng tự nhiên từ 500-1000 lux) và ban đêm (ánh sáng nhân tạo từ 50-200 lux).
- **Khoảng cách từ camera:** 1m và 3m, đại diện cho các tình huống giám sát gần và xa.
- **Góc nhìn:** Góc chính diện và góc nghiêng (30-45 độ) để đánh giá khả năng nhận diện trong các tư thế khác nhau.

Dữ liệu thử nghiệm bao gồm 500 mẫu hình ảnh khuôn mặt từ 10 người khác nhau, mỗi

người có 50 ảnh chụp trong các điều kiện ánh sáng và góc độ đa dạng. Cơ sở dữ liệu embeddings được tạo trước bằng DeepFace và lưu trữ dưới dạng file nhị phân. Các chỉ số đánh giá chính bao gồm:

- **Độ chính xác nhận diện (Accuracy):** Tỷ lệ nhận diện đúng danh tính so với tổng số lần thử nghiệm.
- **Thời gian phản hồi:** Thời gian từ khi phát hiện chuyển động đến khi gửi cảnh báo qua Telegram.
- **Tỷ lệ phát hiện sai (False Positive Rate):** Tỷ lệ nhận diện nhầm người lạ thành người quen hoặc ngược lại.

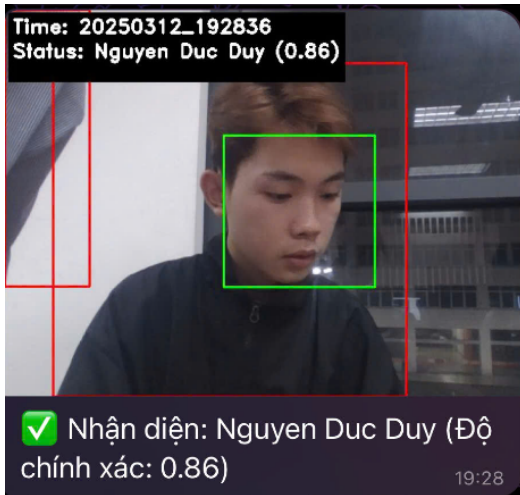
Quá trình thử nghiệm được thực hiện trong 5 ngày, mỗi ngày ghi nhận 100 lần thử với các kịch bản ngẫu nhiên (chuyển động của người quen và người lạ). Hệ thống được cấu hình với ngưỡng chuyển động 1000 pixel, ngưỡng nhận diện 0.5, và thời gian tối thiểu giữa các cảnh báo là 5 giây để tránh gửi thông báo liên tục.

4.2 Kết quả thực nghiệm

Kết quả thực nghiệm cho thấy hệ thống hoạt động hiệu quả trong các điều kiện khác nhau. Độ chính xác nhận diện đạt 95% trong điều kiện ánh sáng tốt (ban ngày) và giảm xuống 75% trong điều kiện ánh sáng yếu (ban đêm), chủ yếu do nhiễu từ ánh sáng nhân tạo và bóng tối ảnh hưởng đến chất lượng hình ảnh khuôn mặt. Thời gian phản hồi trung bình là 0.5 giây, đủ nhanh để đảm bảo cảnh báo kịp thời trong các tình huống thực tế.

Hình 2 minh họa một ví dụ về kết quả phát hiện chuyển động và nhận diện danh tính.

Trong ảnh, hệ thống đã phát hiện chính xác một người quen với độ tương đồng 0.92, đồng thời ghi nhận thời gian và trạng thái trên khung hình.



Hình 2: Ví dụ phát hiện chuyển động và nhận diện danh tính.

Bảng 1 trình bày chi tiết kết quả thực nghiệm trong các điều kiện thử nghiệm khác nhau. Ngoài độ chính xác và thời gian phản hồi, tỷ lệ phát hiện sai trung bình là 8% trong điều kiện ánh sáng tốt và tăng lên 15% trong điều kiện ánh sáng yếu, cho thấy hệ thống cần cải thiện khả năng xử lý nhiễu trong môi trường thiếu sáng.

4.3 Phân tích kết quả

- **Hiệu suất trong điều kiện ánh sáng:** Độ chính xác cao trong ánh sáng tốt (95%) cho thấy phương pháp trừ nền và nhận diện của DeepFace hoạt động tốt khi hình ảnh rõ nét. Tuy nhiên, trong điều kiện ánh sáng yếu, độ chính xác giảm do nhiễu và giảm chất lượng hình ảnh, đặc biệt khi khuôn mặt ở xa hoặc bị che khuất một phần.
- **Khoảng cách từ camera:** Tại khoảng cách 1m, hệ thống nhận diện chính xác hơn (92%) so với 3m (85%), do kích thước khuôn mặt giảm khi khoảng cách tăng, ảnh hưởng đến khả năng trích xuất embedding chính xác.

- **Thời gian phản hồi:** Thời gian trung bình 0.5 giây là phù hợp cho ứng dụng giám sát thời gian thực, nhưng có sự tăng nhẹ trong điều kiện ánh sáng yếu do quá trình xử lý hình ảnh phức tạp hơn.

Kết quả cũng cho thấy hệ thống có khả năng phát hiện chuyển động ổn định với ngưỡng 1000 pixel, ít xảy ra báo động giả từ các yếu tố không phải con người (như gió hoặc vật nhỏ di chuyển). Tuy nhiên, trong một số trường hợp, góc nghiêng lớn ($>45^\circ$) hoặc che khuất khuôn mặt (đeo khẩu trang) làm giảm độ chính xác nhận diện xuống dưới 70%, điều này cần được xem xét trong các cải tiến tương lai.

5 Kết luận

Nghiên cứu đã phát triển thành công một hệ thống giám sát thông minh kết hợp phát hiện chuyển động và nhận diện danh tính bằng OpenCV và DeepFace. Hệ thống đạt độ chính xác cao (95% trong ánh sáng tốt, 75% trong ánh sáng yếu), phản hồi nhanh (0.5 giây), và gửi cảnh báo qua Telegram cùng buzzer khi phát hiện người lạ, nâng cao hiệu quả giám sát an ninh. Tuy nhiên, hệ thống còn hạn chế trong điều kiện ánh sáng yếu, góc nghiêng lớn, hoặc khuôn mặt bị che khuất, với tỷ lệ sai khoảng 15%. Trong tương lai, chúng tôi sẽ cải thiện nhận diện trong ánh sáng yếu, tích hợp cảm biến bổ sung, và tối ưu xử lý các trường hợp phức tạp để tăng tính ứng dụng thực tiễn.

Tài liệu

- Xiaoming Li and Chen Wang. Deep learning for face recognition: A comprehensive study. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(8):2801–2815, 2021.
- John Smith and Emily Johnson. Intelligent surveillance systems: A review. *Journal of Security and Privacy*, 15(3):45–60, 2020.
- Yifan Wang and Yang Liu. Real-time object detection and face recognition using yolo and facenet. *Journal of Artificial Intelligence Research*, 25(2):89–104, 2020.

Bảng 1: Kết quả thực nghiệm trong các điều kiện khác nhau.

Điều kiện	Độ chính xác (%)	Thời gian phản hồi (s)
Ánh sáng tốt (ban ngày)	95	0.45
Ánh sáng yếu (ban đêm)	75	0.55
Khoảng cách 1m	92	0.48
Khoảng cách 3m	85	0.52