

THIẾT KẾ HỆ THỐNG TÁCH VÀ NHẬN DẠNG KHUÔN MẶT TỪ VIDEO DESIGNING A VIDEO-BASED FACE DETECTION AND RECOGNITION SYSTEM

Đậu Trọng Hiền, Ngô Quốc Cường, Trần Tùng Giang

Trường Đại Học Sư Phạm Kỹ Thuật TP. HCM

TÓM TẮT

Nhận dạng mặt người trong video là đề tài đang được chú ý trong những năm gần đây trong lĩnh vực thị giác máy tính. So với nhận dạng trên ảnh tĩnh truyền thống, nhận dạng trên video có nhiều lợi thế về lượng thông tin để tăng độ chính xác và ổn định. Tuy nhiên việc nhận dạng ảnh trong video cũng gặp nhiều khó khăn vì sự biến thiên lớn về tỉ lệ khuôn ảnh, chất lượng ảnh trong video kém, độ chói và tư thế khuôn mặt thay đổi, khuôn mặt bị che phủ. Bài báo này đưa ra phương pháp nhận dạng khuôn mặt trong video dựa vào ảnh của các khung hình trong video. Quá trình nhận dạng khuôn mặt trong video gồm ba công đoạn chính-Tách khuôn mặt dùng thuật toán Haar-like, trích đặc trưng khuôn mặt dùng PCA và nhận dạng dùng mạng nơ ron đa lớp. Thuật toán nhận dạng trong bài này được kiểm tra bằng đoạn phim có 1000 khung hình. Độ chính xác của thuật toán nhận dạng là 98%.

Từ khóa: Khung hình trong video, đặc tính Haar-Like, nhận dạng khuôn mặt, phân tích thành phần chính, mạng nơ ron, vector riêng, khuôn mặt riêng.

ABSTRACT

Face recognition in videos has been a hot topic in computer vision in recent years. Compared to traditional face analysis, video-based face recognition has the advantages of more abundant information to improve accuracy and robustness, but also suffers from large scale variations, low quality of facial images, illumination changes, pose variations and occlusions. The paper presents a method for face recognition based on video-image based methods. The proposed method consists of three stages: face detection using Haar-Like feature, feature extraction using principle component analysis, and recognition using the feed forward back propagation Neural Network. The algorithm has been tested on a video with 1000 frames (1000 images). Test results gave a recognition rate of 98%.

Key words: Video frame, Haar-Like feature, Face recognition, Principal component analysis (PCA), Artificial Neural network (ANN), Eigenvector, Eigenface.

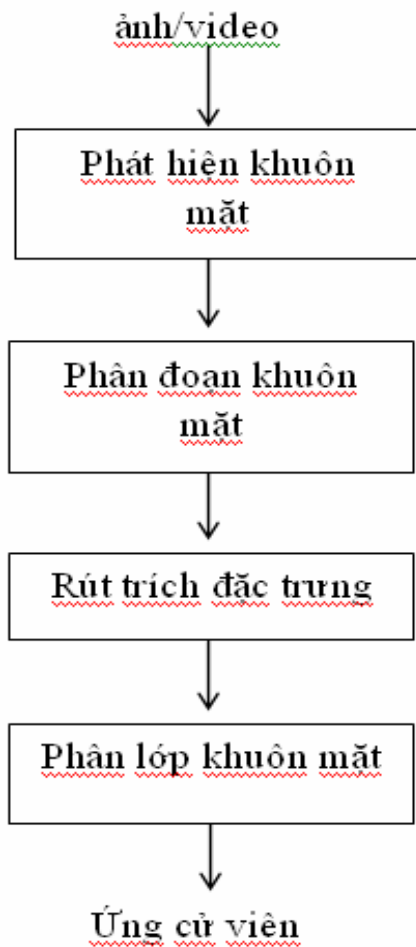
I. Giới thiệu

Một hệ thống nhận dạng mặt người thông thường bao gồm bốn bước xử lý sau: phát hiện khuôn mặt (face detection), phân đoạn khuôn mặt (face alignment hay segmentation), rút trích đặc trưng (feature extraction), và phân lớp khuôn mặt (face classification). *Phát hiện khuôn mặt:* đi tìm và định vị những vị trí khuôn mặt xuất hiện trong ảnh hoặc trên các frame

video. *Phân đoạn khuôn mặt* sẽ xác định vị trí mắt mũi, miệng, và các thành phần khác của khuôn mặt và chuyển kết quả này cho bước *rút trích đặc trưng*. Từ những thông tin về các thành phần trên khuôn mặt, chúng ta có thể dễ dàng tính được véc-tơ đặc trưng trong bước rút trích đặc trưng. Những véc-tơ đặc trưng này sẽ là dữ liệu đầu vào cho một mô hình đã được huấn luyện trước để *phân loại khuôn mặt*. Bên

cạnh những bước chính nêu trên, chúng ta còn có thể áp dụng thêm một số bước khác như *tiền xử lý*, *hậu xử lý* nhằm làm tăng độ chính xác cho hệ thống.

Do một số thông số như: tư thế khuôn mặt, độ sáng, điều kiện ánh sáng, v.v..., *phát hiện khuôn mặt* được đánh giá là bước khó khăn và quan trọng nhất so với các bước còn lại của hệ thống.

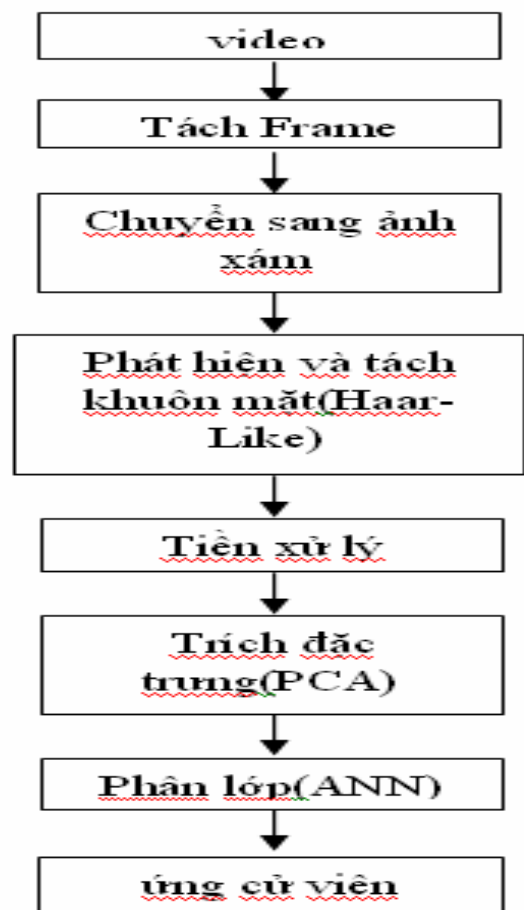


Hình 1: Các bước chính trong một hệ thống nhận dạng mặt người

II. Thuật toán đề nghị :

Trong đề tài này nhóm tác giả sử dụng thuật toán Haar-Like để tách khuôn mặt ra khỏi khuôn hình. Haar-Like là thuật toán ổn định và được sử dụng phổ biến thông qua OPEN CV hoặc thư viện liên kết động DLL chính vì vậy việc viết chương trình sử dụng thuật

toán này rất thuận tiện. Đối với giai đoạn trích đặc trưng và phân lớp khuôn mặt ta có thể dùng PCA hoặc nơ ron riêng rẽ. Tuy nhiên nếu ta chỉ dùng PCA và dựa trên khoảng cách Euclide để xác định ứng cử viên thì độ chính xác trong việc nhận dạng không được cao. Nếu ta dùng mạng nơ ron để nhận dạng thì độ chính xác khá cao[5] tuy nhiên khi tập mẫu lớn thì ta phải tăng số nút mạng để tăng năng lực học của mạng dẫn đến tốc độ nhận dạng của mạng sẽ giảm. Chính vì vậy trong đề tài này nhóm tác giả sử dụng PCA để trích đặc trưng còn mạng nơ ron sẽ đóng vai trò phân lớp khuôn mặt. Bằng việc sử dụng PCA để trích đặc trưng ta đã giảm đáng kể số lượng mẫu học cho mạng nơ ron.



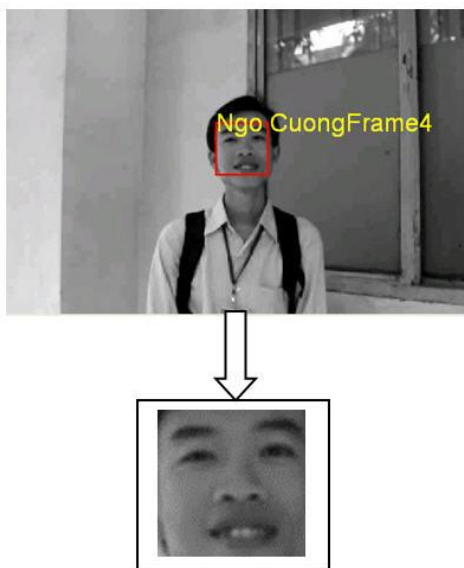
Hình 2: Sơ đồ chi tiết hệ thống nhận dạng khuôn mặt

Tách Frame:

Video từ camera sẽ được tách thành từng khung hình(frame) ta có thể xem mỗi khung hình là một ảnh tĩnh để xử lý.

Phát hiện và tách khuôn mặt(Haar-Like)

Khuôn mặt sẽ được theo dõi, phát hiện và tách ra khỏi từng frame hình dùng thuật toán Haar-Like[6].



Hình 3: Tách khuôn mặt dùng Haar-Like.

Tiền xử lý

Ảnh khuôn mặt sau khi được tách được đưa vào chuẩn hóa kích thước về kích thước cho trước $X \times Y$ pixels (trong đề tài này tác giả sử dụng kích thước 92x92 pixels). Sau đó ảnh được cân bằng histogram để tăng hiệu quả nhận dạng. Trong giai đoạn huấn luyện ảnh qua khâu tiền xử lý sẽ được đưa vào thư viện ảnh mẫu.

Trích đặc trưng(PCA)

Đầu tiên ta sẽ đi tính tập khuôn mặt riêng từ thư viện ảnh mẫu. Mỗi một khuôn mặt có thể đại diện bằng việc kết hợp tuyến

tính của các khuôn mặt riêng. Và ta có thể tính gần đúng bằng cách chọn M' khuôn mặt riêng tốt nhất tức là những khuôn mặt riêng có giá trị riêng lớn nhất. Nó miêu tả sự biến thiên lớn nhất trong tập ảnh mẫu. M' khuôn mặt riêng này tạo ra một không gian con M' chiều gọi là “không gian khuôn mặt”. của tất cả các khuôn mặt có thể tồn tại. Để tính tập khuôn mặt riêng này ta sử dụng thuật toán PCA trong [1],[2]. Giả sử ta có thư viện ảnh $\Gamma_1, \Gamma_2, \dots, \Gamma_m$ với mỗi ảnh là $I(x,y)$. Chuyển mỗi ảnh thành một vector ta sẽ có ma trận thư viện ảnh ($m \times p$) trong đó $p = x \times y$. Trình tự tính PCA[3]:

Tính khuôn mặt trung bình:

$$\Psi = \frac{1}{m} \sum_{i=1}^m \Gamma_i \quad (1)$$

Trừ mỗi ảnh cho mặt trung bình:

$$\Phi_i = \Gamma_i - \Psi, i = 1, 2, \dots, m \quad (2)$$

Đặt: $A = [\Phi_1, \Phi_2, \dots, \Phi_m]$ ta được

ma trận hiệu có kích thước $m \times p$.

Xây dựng ma trận covariance:

$$C_{mm} = A_{mp} \times A_{pm}^T \quad (3)$$

Tính vector riêng V_{mm} và trị riêng λ_m của ma trận C sử dụng phương pháp Jacobi và sắp xếp vector riêng theo thứ tự giá trị riêng lớn nhất.

Tính khuôn mặt riêng:

$$U_k = \sum_{n=1}^m \Phi_n V_{kn}, k = 1, 2, \dots, m \quad (4)$$

Thay vì sử dụng m khuôn mặt riêng ta có thể sử dụng m' khuôn mặt riêng với $m' < m$.

Dựa vào khuôn mặt riêng ta tính vector khuôn mặt bằng phương trình:

$$W_k = U_k^T(\Gamma - \Psi), k = 1, 2, \dots, m' \quad (5)$$

Từ đó tạo ra **vector đặc trưng**:

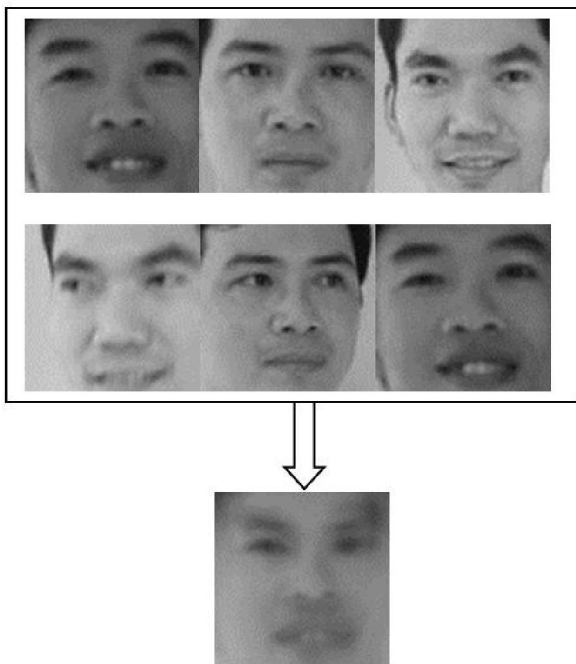
$$\Omega^T = [w_1, w_2, \dots, w_{m'}]$$

Ta có thể tái tạo lại khuôn mặt bằng phương trình:

$$\Gamma' = \Psi + \Phi_f \quad (6)$$

trong đó $\Phi_f = \sum_{i=1}^{m'} w_i U_i$

Trong đề tài này nhóm tác giả chọn 3 người làm mẫu với số lượng ảnh mỗi người là 40 với các tư thế chính diện, trái, phải, ngược lên trên, xuống dưới khác nhau. Như vậy tổng số ảnh là 120. Số khuôn mặt riêng ta có thể chọn sao cho ít nhất nhưng vẫn đảm bảo được độ chính xác trong quá trình nhận dạng.

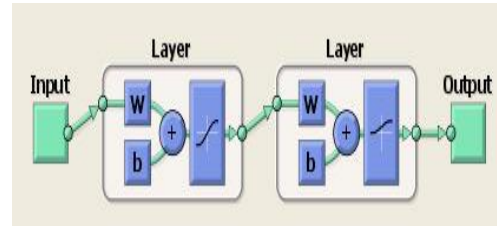


Hình 4: Ảnh các khuôn mặt và ảnh trung bình

Huấn luyện mạng nơ ron

Tập ảnh mẫu sau khi qua khâu trích đặc trưng PCA tạo ra vector đặc trưng Ω^T được

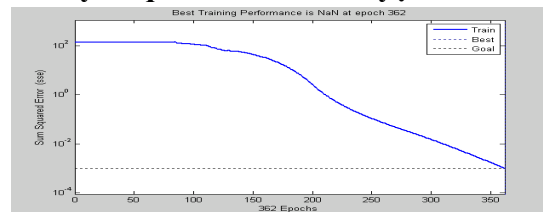
đưa vào ngõ vào của mạng nơ ron để huấn luyện. Trong đề tài này tác giả dùng mạng nơ ron MLP 1 lớp ẩn. Ngõ ra có 3 nút đại diện cho 3 người.



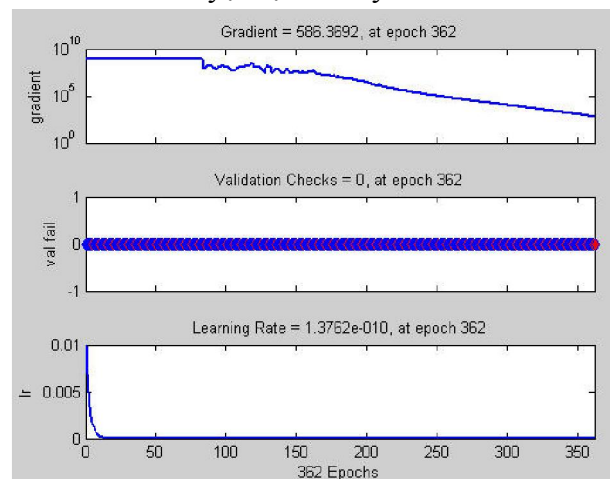
Hình 5: Kiến trúc mạng nơ ron
 Hàm kích hoạt cho ngõ ra của lớp vào là tansig và lớp ra là logsig.

Thuật toán học theo phương pháp thích nghi.

Sau đây là quá trình huấn luyện:



Hình 6: Tổng sai số bình phương trong quá trình huấn luyện tại chu kỳ thứ 362



Hình 7: Quá trình huấn luyện

Chạy thử nghiệm chương trình với số lượng ảnh trong tập mẫu là 120, video kiểm tra có số frame là 1000 với các góc mặt quay khác nhau và camera rung động khi thu hình.

Số mặt riêng	Số lượng ảnh nhận dạng đúng trong tập mẫu	Số lượng frame nhận dạng đúng trong video kiểm tra	Độ chính xác trên tập mẫu (%)	Độ chính xác trên tập kiểm tra (%)
30	120	951	100	95.1
60	120	983	100	98.3
80	120	973	100	97.3
120	120	964	100	96.4

Bảng 1: Kết quả thử nghiệm.

So sánh với thuật toán K-means, Fuzzy Ant với fuzzy C-means.[4]:

Phương pháp	tỉ lệ nhận dạng
K-means	86.75
Fuzzy Ant with fuzzy C-means	94.82
Phương pháp đề nghị trong bài	98.3

Bảng 2: Bảng so sánh kết quả.

TÀI LIỆU THAM KHẢO

[1]. Kirby, M., and Sirovich, L., "Application of the Karhunen-Loeve procedure for the characterization of

IV. Kết luận

Thuật toán nhận dạng cho phép nhận dạng khuôn mặt với độ chính xác cao với góc nhìn nghiêng không quá 30 độ. Tốc độ của thuật toán là khá nhanh. Có thể nhận dạng các video bị rung do không dùng chân máy quay. Thuật toán nhận dạng khuôn mặt trong video kết hợp giữa phương pháp phân tích thành phần chính(PCA) và mạng nơ ron đem lại kết quả khá cao. Kết quả được so sánh với thuật toán K-means và Fuzzy với fuzzy C-means và kết quả cho thấy thuật toán được đề nghị trong bài có độ chính xác cao hơn. Trong bảng 1 chúng ta có thể thấy độ chính xác thay đổi tùy theo số khuôn mặt riêng và độ chính xác đạt cao nhất là 98.3%. Chỉ những khuôn mặt riêng có giá trị riêng cao nhất mới thực sự cần thiết cho việc nhận dạng. Như ta thấy trong bảng 1 với M=50 thì độ chính xác đạt cao nhất.

Bằng cách chọn PCA để trích đặc trưng ta đã giảm số chiều của không gian mẫu chỉ còn bằng số mặt riêng tương ứng với trị riêng cao nhất.

human faces", *IEEE PAMI*, Vol. 12, pp. 103-108, (1990).

[2]. S. Gong, S. J. McKeANNA, and A. Psarron, *Dynamic Vision*, Imperial College Press, London, 2000.

- [3]. Zhujie, Y.L.Y., 1994. Face recognition with eigenfaces. Proc. IEEE Intl. Conf. Industrial Technol., pp: 434-438 .
- [4] S. Makdee, C. Kimpan, S. Pansang, “Invariant range image multi-pose face recognition using Fuzzy ant algorithm and membership matching score.” Proceedings of 2007 IEEE International Symposium on Signal Processing and Information Technology, 2007, pp. 252-256.
- [5]. Đậu Trọng Hiền, “Nghiên cứu thiết kế thuật toán nhận dạng khuôn mặt dựa trên mạng nơ ron,” *Tạp chí khoa học kỹ thuật* 2010.
- [6] T. Sakai, M. Nagao, and S. Fujibayashi, “Line Extraction and Pattern Detection in a Photograph”, *Pattern Recognition*, vol. 1, pp. 233-248, 1969.