

MỘT KỸ THUẬT NHẬN DẠNG BIỂU CẢM KHUÔN MẶT DỰA TRÊN MÔ HÌNH CHẤT LIỆU

Huỳnh Cao Tuấn*, Đỗ Năng Toàn*, Nguyễn Thanh Bình#, Lâm Thành Hiền*

* Viện Công nghệ Thông tin – Đại học Quốc gia Hà Nội

Học viện Công nghệ Bưu chính viễn thông

* Trường Đại học Lạc Hồng

Tóm tắt: Kỹ thuật nhận dạng khuôn mặt phát hiện cảm xúc của người dựa trên các biểu hiện trên khuôn mặt của họ (vui, buồn, ghê tởm, bất ngờ, tức giận, biểu cảm sợ hãi) là kỹ thuật phát hiện đối tượng nâng cao. Điều này đã tìm thấy các ứng dụng trong các lĩnh vực kỹ thuật như Human-computer-Interaction (HCI) và giám sát an ninh. Nó thường yêu cầu xử lý và ra quyết định nhanh. Bài báo này trình bày một kỹ thuật phát hiện khuôn mặt dựa trên mô hình chất liệu và nhận dạng cảm xúc.

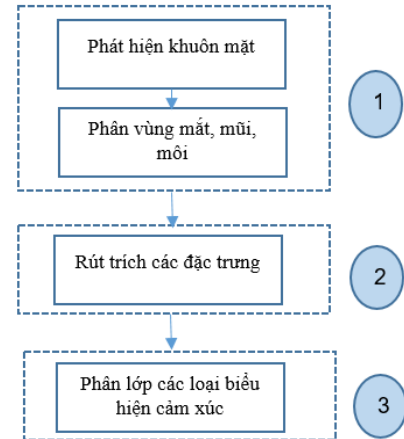
Từ khóa: facical expression recognition, biometric recognition, emotion recognition

I. GIỚI THIỆU

Việc nhận biết biểu hiện khuôn mặt là để phát hiện cảm xúc của con người dựa trên biểu hiện. Biểu cảm trên khuôn mặt hiện đang được nghiên cứu và ứng dụng trên nhiều lĩnh vực khác nhau như: hệ thống xác lái xe an toàn với mục tiêu tự động phát hiện tình trạng mệt mỏi của người lái xe làm giảm khả năng đề phòng rủi ro của người lái xe, đồng thời hệ thống còn hỗ trợ đưa ra những biện pháp phù hợp để phòng tránh tai nạn[1], ứng dụng trong lĩnh vực y học với việc chẩn đoán các biểu hiện tâm lý thông qua biểu cảm trên gương mặt, Các nghiên cứu liên quan đến trạng thái cảm xúc như khoa học hành vi, tâm thần học, thần kinh học: nhằm cải tiến việc xử lý dữ liệu trạng thái cảm xúc hiệu quả hơn, độ chính xác cao hơn[2,3], Giáo dục: các hệ thống giảng dạy có thể nhận dạng cảm xúc và trạng thái tiếp thu của học sinh.

Bài toán này bao gồm ba bước chính: phát hiện khuôn mặt người có trong ảnh dựa trên mô hình chất liệu, rút trích đặc trưng và phân loại biểu hiện cảm xúc. Số lượng nghiên cứu được thực hiện trong mỗi bước này là khá đáng kể và đáng chú ý.

Phần còn lại của bài báo được trình bày như sau: Nội dung được giải thích trong phần II. Tổng kết đánh giá thực nghiệm được trình bày trong phần III. Các nhận xét kết luận được đưa ra trong phần IV.

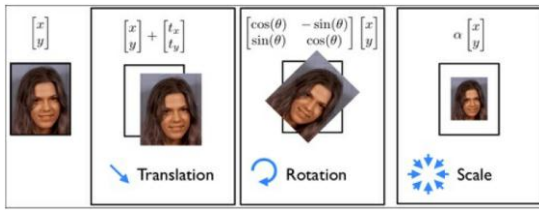


Hình 1. Biểu đồ luồng nhận dạng biểu cảm

II. TRÍCH RÚT ĐẶC TRƯNG BIỂU CẢM KHUÔN MẶT

1) Mô hình hình dạng (Active Appearance Models)

AAM (active appearance model) là một thuật toán phổ biến trong lĩnh vực thị giác máy nhằm mục tiêu tối ưu một mô hình thống kê hình ảnh thể hiện của đối tượng vào một ảnh đầu vào mới, kết quả của quá trình tối ưu là một bộ điểm điều khiển thể hiện cấu trúc của đối tượng đã được học với các tọa độ tương ứng với thể hiện trong ảnh đầu vào của đối tượng cùng với nó là một bộ các tham số mô hình thống kê đã được ước lượng mà từ những tham số mô hình này, ta có thể dễ dàng tái cấu trúc cả về hình dạng cũng như kết cấu hình ảnh của đối tượng tương ứng một cách tương đối với thể hiện của đối tượng ở trong ảnh thử nghiệm. Trong AAM, đối tượng quan tâm được mô hình hóa bởi một tập điểm mô tả hình dạng và kết cấu hình ảnh của nó, kết cấu hình ảnh được lấy chính là mẫu của những giá trị cường độ ảnh trong những vùng được giới hạn bởi tập điểm điều khiển. Mô hình thống kê của đối tượng đảm bảo có thể mô tả được những biến thể về hình dạng và những biến thể về kết cấu hình ảnh cũng như mối tương quan giữa chúng. Các vấn đề nổi bật được quan tâm trong phương pháp này là việc xây dựng mô hình thống kê cho đối tượng ảnh và việc thiết kế thuật toán tối ưu để thực hiện tìm kiếm. Xây dựng mô hình thống kê toán học cho đối tượng bao gồm việc xây dựng mô hình mô hình thống kê toán học cho hình dạng, mô hình mô hình thống kê toán học cho kết cấu hình ảnh và việc kết hợp hai mô hình đó để ra được mô hình cho đối tượng.



Hình 2. Phép biến đổi Similarity

Mô hình hình dạng của đối tượng được biểu diễn bởi một tập hợp có thứ tự các điểm điều khiển. Một vấn đề quan trọng trong xây dựng mô hình chính là việc cần thiết phải lựa chọn được ra những điểm điều khiển phù hợp với loại đối tượng cần được mô hình hóa. Khi xây dựng tập dữ liệu huấn luyện, tập dữ liệu bao gồm một tập ảnh của đối tượng cùng với thông tin đánh dấu tập điểm điều khiển mô tả thể hiện của đối tượng được quan tâm tương ứng trong các ảnh. Đối tượng hình học chính là tập hợp có đánh thứ tự của các điểm điều khiển mô tả hình dạng của đối tượng ảnh được quan tâm, được biểu diễn bởi n điểm trong không gian d chiều, được thể hiện bằng một vector nd chiều với các giá trị trong vector chính giá trị tọa độ cụ thể trên từng trục không gian của từng điểm. Ví dụ, trong ảnh 2 chiều, ta có thể mô tả n điểm điều khiển, $\{(x_i, y_i)\}$, thành vector $2n$ thành phần: $x = (x_1, \dots, x_n, y_1, \dots, y_n)^T$.

Sau khi thể hiện lại các đối tượng hình dạng thành các vector tương ứng, thực hiện mô hình hóa trên dữ liệu là tập các vector này. Bước đầu là thực hiện chuẩn hóa các đối tượng hình dạng này vào chung một không gian tọa độ nhằm chuẩn hóa các đối tượng hình dạng để tổng các khoảng cách của mỗi đối tượng hình dạng tới kỳ vọng của chúng

$$D = \left(\sum_{i=1}^n |x_i - \bar{x}|^2 \right)$$

đạt cực tiểu dựa trên PCA.

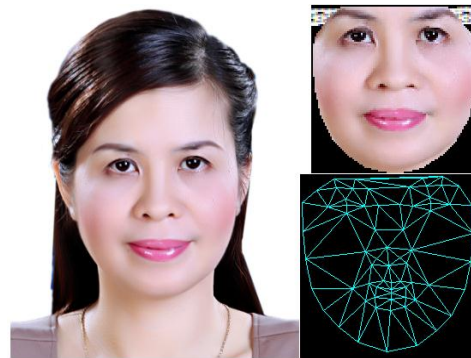
Sau khi đã mô hình hóa được các đối tượng hình dạng, thực hiện tiếp quá trình mô hình hóa kết cấu hình ảnh của các đối tượng được quan tâm trong ảnh. Kết cấu hình ảnh của đối tượng được hiểu là các giá trị cường độ ảnh được giới hạn trong vùng ảnh được bao bởi đối tượng hình dạng tương ứng. Dữ liệu đối tượng hình dạng và kết cấu hình ảnh của đối tượng được quan tâm trong các mẫu có thể được biểu diễn dưới dạng các tham số. Vấn đề tiếp theo là thực hiện mô hình hóa tổng quát cho các dữ liệu này và tương tự phương pháp được lựa chọn là phân tích thành phần chính PCA.

Giải thuật tìm kiếm tối ưu được sử dụng trong AAM được thiết kế cho phép tự động ước lượng các tham số mô hình cái mà có thể tổng hợp ra được ảnh mẫu gần nhất có thể với ảnh mục tiêu đầu vào nhằm mục đích cực tiểu sự sai lệch giữa ảnh mẫu đầu vào và ảnh tổng hợp được theo mô hình đã được lựa chọn.[4]

2) Mô hình kết cấu hình ảnh

Sau khi đã mô hình hóa được các đối tượng hình dạng, ta thực hiện tiếp quá trình mô hình hóa kết cấu hình ảnh của các đối tượng được quan tâm trong ảnh. Ở đây, kết cấu hình ảnh của đối tượng được hiểu là các giá trị cường độ ảnh được giới hạn trong vùng ảnh được bao bởi đối tượng hình dạng tương ứng

Quá trình lấy ra được dữ liệu kết cấu hình ảnh này được các tác giả thực hiện qua hai bước. Đầu tiên là thực hiện quá trình tam giác hóa trên tập điểm điều khiển để lấy ra được một tập các vùng ảnh và bước tiếp theo là thực hiện quá trình biến đổi hình học từng phần trên ảnh tương ứng với từng tam giác đã được tính toán và ghi lại trên một đối tượng hình học tiêu chuẩn đã được chọn trước cụ thể là theo phép biến đổi affine. Phép biến đổi này được gọi là Piece-wise Affine.



Hình 3. Đối tượng hình dạng và kết cấu hình ảnh

Mỗi mẫu kết cấu hình ảnh lấy được đều có chứa ít nhiều những biến dạng gây ra bởi sự chiếu sáng khi thu nhận ảnh.

3) Mô hình chất liệu khuôn mặt

Mô hình chất liệu mặt nhằm ràng buộc tập các điểm điều khiển trong không gian khuôn mặt kể cả khi điều kiện ánh sáng thay đổi và nền phức tạp. Chất liệu ảnh mặt có thể được mô hình hóa dựa vào bề mặt của chất liệu da và cấu trúc hình học của vùng mặt. Trong bài báo này, chất liệu khuôn mặt được quan tâm có một số tính chất như bao gồm chất liệu da cục bộ, đặc trưng thể hiện cho vùng da mặt của khuôn mặt trong ảnh, và phạm vi khuôn mặt được giới hạn tương đối trong một hình ellipse bao quanh vị trí khuôn mặt.[4]

Vấn đề đầu tiên ta quan tâm đến là việc xây dựng chất liệu da cục bộ và tách da. Da là một thành phần cấu tạo nên chất liệu mặt, do đó để xây dựng mô hình chất liệu mặt thì vấn đề quan trọng là phải xác định được các pixel là da và thực hiện phân cụm các pixel thành nhóm da và nhóm không da. Trong bước này ta sẽ phải duyệt qua các điểm trong ảnh vào. Tại mỗi điểm, ta sẽ phải xác định xem nó có phải là màu da hay không? Nếu là màu da, ta sẽ gán giá trị 1 tại điểm đó bằng 1, nếu không ta sẽ gán giá trị 0. Kết quả ta có 1 ảnh nhị phân, với các điểm 1 (điểm trắng) là các điểm màu da. Trong nhiều nghiên cứu, các tác giả đã cố gắng xây dựng một mô hình chất liệu da tổng quát, ví dụ như C. Garcia, G. Zikos, G. Tziritis

[5] đưa ra công thức được xây dựng trên hai hệ màu YCrCb và HSV hay K. Sandeep và A.N. Rajagopalan [6] xây dựng mô hình chất liệu da với histogram được tổng hợp trên một tập ảnh màu da (tập ảnh mẫu được chuẩn bị trước), nhưng không phải trên hệ màu RGB mà là trên hệ màu HSV và trên 2 thành phần H và S (histogram 2 chiều).

Chúng tôi quan tâm đến việc xây dựng mô hình chất liệu da cục bộ, tức là công thức được tạo ra với riêng khuôn mặt đang quan tâm trong ảnh. Đây là cơ sở để có thể áp dụng được cho những trường hợp khuôn mặt được thu nhận trong những điều kiện chiếu sáng và chủng tộc khác nhau. Một cách tổng quát, mô hình chất liệu da là $\{C_i, r_i\}, i = 1..n$ trong đó C_i, r_i tương ứng là tâm và bán kính của vùng giá trị trên trục tương ứng kênh màu i , n là tổng số kênh màu.

Việc tách ngưỡng ở đây được thực hiện dựa trên ý tưởng của một hình ellipsoid trong không gian n chiều, các tham số của hình ellipsoid về tâm và bán kính trên các trục được sử dụng tương ứng với các tham số của mô hình chất liệu da. Một điểm không phải chất liệu da nếu nó nằm ngoài hình ellipsoid và là chất liệu da trong trường hợp ngược lại. Trên cơ sở ý tưởng đó, thuật toán tách ngưỡng ellipsoid được mô tả cụ thể như sau:

Input: Ảnh I có giá trị màu n và chất liệu da

Output: Ảnh mặt nạ M kết quả phân ngưỡng

Các bước:

Với mỗi điểm $P = (p_1, p_2, \dots, p_n)$

có tọa độ (x, y) là một điểm ảnh của I :

$d := 0$;

for $i := 1$ to n

$d := d + ((p_i - C_i)/r_i)^2$;

if $(d \leq 1)$

$M(x, y) := 1$;

Else

$M(x, y) := 0$;

Một vấn đề trọng tâm cần được giải quyết chính là việc xây dựng mô hình chất liệu da cục bộ. Nói cách khác, đó chính là vấn đề xây dựng thuật toán ước lượng các tham số cho mô hình chất liệu da. Việc ước lượng các tham số này được áp dụng với đầu vào là một ảnh khuôn mặt với ý tưởng là tìm một cụm các điểm ảnh gần nhau nhất và chiếm đa số các điểm ảnh của ảnh khuôn mặt. Do đầu vào là ảnh khuôn mặt nên kết quả của cụm này sẽ tương ứng với vùng ảnh da do nó chiếm đa số trong ảnh khuôn mặt. Giải thuật được thiết kế với quá trình lặp tương tự như ý tưởng trong thuật toán gom cụm k-means. Trên cơ sở ý tưởng đó, thuật toán ước lượng chất liệu da dựa trên tỉ lệ được mô tả cụ thể như sau:

Input: Ảnh I kích thước $w \times h$ chiều và tỉ lệ

$r(0.5 < r < 1.0)$

Output: Chất liệu da

Các bước:

Khởi tạo

C là màu trung bình của I

L là mảng số thực có kích thước $w \times h$

$m := r * w * h$

(1)

Repeat

Với mỗi điểm ảnh của $I(x, y)$:

$L(y * w + x) := \text{distance}(C, I(x, y))$

(2)

Sắp xếp L tăng dần

$C :=$ trung bình của m điểm ảnh tương ứng m giá trị đầu tiên trong L

Until (C không đổi)

Dùng m điểm ảnh tương ứng m giá trị đầu tiên trong ảnh L , tính các giá trị C_i và r_i tương ứng

Với mô hình các bước mô tả như trên, với một ảnh khuôn mặt đầu vào, ta có thể ước lượng được tham số cho mô hình chất liệu da mặt và sử dụng nó để tách ngưỡng lại ảnh ban đầu, kết quả ta sẽ thu được ảnh nhị phân thể hiện các vùng da. Thêm vào đó để có thể thực hiện tách vùng khuôn mặt, chúng tôi thực hiện ước lượng thêm một vùng ellipse quanh vị trí ban đầu. Như vậy, vùng giao sẽ là kết quả vùng tương ứng với vùng ảnh khuôn mặt đầu vào. Tuy nhiên do bản chất của việc tách ngưỡng với mô hình chất liệu da cục bộ ở bước trên, vùng khuôn mặt sẽ xuất hiện một số lỗ thủng tương ứng một số vị trí có chất liệu khác biệt với da như vùng mắt, miệng, lông mày. Do đó, cần thực hiện một thao tác lấp lỗ hổng để lấy được vùng mặt quan tâm.

4) Thuật toán xác định vùng khuôn mặt

Trên cơ sở mô hình chất liệu khuôn mặt đã được mô tả ở trên, chúng tôi đưa ra thuật toán xác định vị trí ảnh khuôn mặt trên ảnh tổng quát. Thuật toán thực hiện quá trình ước lượng chất liệu mặt và định vị với từng khuôn mặt trong ảnh. Trên cơ sở những kết quả đó, đầu ra của thuật toán là ảnh nhị phân đã được tổng hợp lại. Sau đây là sơ đồ của thuật toán xác định chính xác vùng mặt dựa trên mô hình chất liệu da đã xây dựng được: [4]

Input: Ảnh I

Output: Ảnh mặt nạ M thể hiện vùng khuôn mặt

Các bước:

Khởi tạo

Faces là danh sách các hình chữ nhật

Các giá trị của M gán bằng 0

Phát hiện các khuôn mặt, lưu vào faces

Với vị trí faces[i]:

Chọn R là vùng ảnh mờ rộng của faces[i]

Khởi tạo ảnh I' là ảnh n chiều tương ứng vùng của I

Tính chất liệu da bằng thuật toán ước lượng chất liệu da dựa trên tỉ lệ

Tính mặt nạ M' bằng thuật toán tách ngưỡng ellipsoid

Ước lượng E là mặt nạ ellipse tương ứng của R

$M' := M' \text{ AND } E$

Lấp các lỗ trong mặt nạ M'

Cập nhật M' vào M

A. Phương pháp rút trích đặc trưng

Giai đoạn rút trích đặc trưng là giai đoạn quan trọng trong hầu hết các hệ thống nhận dạng, xử lý. Trích rút đặc trưng là kỹ thuật sử dụng các thuật toán để lấy ra những thông tin mang những đặc điểm riêng biệt của một người.

Các khâu trong quá trình trích chọn đặc trưng:

- Đầu vào: Ảnh đã được chuẩn hóa.
- Đầu ra: vector đặc trưng của ảnh đầu vào.

1) Eigenfaces

Eigenface đã được M. A. Turk và A. P. Pentland đề xuất năm 1991. Ý tưởng chính của eigenface là để lấy được các đặc trưng trong nghĩa toán học thay vì đặc trưng về mặt vật lý bằng cách sử dụng chuyển đổi toán học để được nhận dạng. [7]

Giai đoạn đầu tiên, một số lượng lớn các hình ảnh khuôn mặt được đưa vào tập huấn luyện. Những hình ảnh được huấn luyện là đại diện tốt nhất cho tất cả các khuôn mặt có thể được gặp phải. Kích thước, hướng và cường độ ánh sáng nên được chuẩn hóa. [7]



Hình 4. Tập ảnh được dùng để huấn luyện.

Bước 1: Chuẩn hóa ma trận ảnh

Cho vector $X = \{x_1, x_2, \dots, x_N\}$, trung bình và chuẩn X được tính như sau:

$$\text{mean}(X) = \bar{X} = \frac{1}{N} \sum_{i=1}^N x_i \quad (3)$$

$$\text{std}(X) = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{X})^2}{N-1}} \quad (4)$$

Theo từ một ảnh Γ ban đầu, ta có trung bình $\bar{\Gamma}$ và chuẩn $\text{std}(\Gamma)$ của Γ . Phép chuẩn hóa ma trận ảnh được thực hiện theo công thức sau:

$$\Gamma'(x, y) = \frac{\Gamma(x, y) - \bar{\Gamma}}{\text{std}(\Gamma)} \quad (5)$$

Bước 2: Tính trung bình của tập ảnh đầu vào. Bước này cho ta xác định được gốc tọa độ không gian vector biểu diễn các ảnh. Ảnh trung bình được tính bằng trung bình cộng của tất cả các ảnh thể hiện sự giống nhau nhất giữa tất cả các khuôn mặt.

Gọi số ảnh học là M , mỗi ảnh có kích thước $R \times C$.

$$S = \{\Gamma_1, \Gamma_2, \dots, \Gamma_M\} \quad (6)$$

Trung bình ảnh M trên được tính theo công thức:

$$\Psi = \frac{1}{M} \sum_{i=1}^M \Gamma_i \quad (7)$$

Bước 3: Tính eigenface

Eigenface chính là các vector riêng tương ứng với các giá trị riêng v có giá trị lớn nhất của ma trận hiệp phương sai - covariance:

$$C = \frac{1}{M} \sum_{i=1}^M \Phi_i \Phi_i^T = A \cdot A^T \quad (8)$$

Để tính nhanh các eigenface này, người ta chuyển sang tính toán với ma trận $L = A^T \cdot A$. Ma trận C có kích thước $N_2 \times N_2$, vấn đề tìm vector riêng u_i của ma trận C khó thực hiện vì kích thước quá lớn. Để tìm vector riêng u_i của C ta thực hiện như sau:

+ Giả sử v_i là vector riêng của $A^T A$, tức là:

$$A^T A v_i = \lambda_i v_i \quad (9)$$

Nhân 2 vế với ma trận A ta được

$$A A^T A v_i = \lambda_i A v_i \quad (10)$$

Vậy $A v_i$ là vector riêng của C . Bây giờ ta cần tìm vector riêng v_i của ma trận $L = A^T A$, ma trận L có kích thước $M \times M$.

Giả sử v_i là vector riêng tìm được của L . Lúc đó, vector riêng u_i của ma trận C cần tìm là:

$$u_i = \sum_k^M V_{ik} \Phi_k, \quad i = \overline{1, M} \quad (11)$$

Ưu điểm:[8]

- Thuật toán đơn giản, dễ cài đặt.
- Thời gian lưu trữ và xử lý hiệu quả.
- PCA làm giảm kích thước của một hình ảnh rất nhiều trong một thời gian ngắn.
- Tính chính xác trên 90% với khuôn mặt phía chính diện.

Nhược điểm:[8]

- Nhạy cảm với điều kiện ánh sáng và vị trí chụp
- Có mối tương quan cao giữa dữ liệu training và dữ liệu nhận dạng
- Tìm các eigenvectors và eigenvalues tốn thời gian
- Kích thước và vị trí của mỗi ảnh khuôn mặt phải vẫn tương tự PCA (Eigenface) tiếp cận các đặc trưng bản đồ không gian con.

2) Fisherfaces

Phương pháp nhận dạng khuôn mặt fisherface được mô tả bởi Belhumeur và cộng sự sử dụng cả phân tích thành phần cơ bản - Principal Components Analysis (PCA) và phân tích phân biệt tuyến tính - Linear Discriminant Analysis (LDA) để tạo ra một ma trận chiếu không gian con, tương tự như được sử dụng trong phương pháp eigenface. Tuy nhiên, phương pháp fisherface có thể tận dụng lợi thế của thông tin trong các lớp, giảm thiểu sự thay đổi trong mỗi lớp, nhưng vẫn tối đa hóa tách lớp. Giống như quá trình xây dựng eigenface, bước đầu tiên là lấy từng mảng hình ảnh ($n \times m$) và định hình lại thành một $((N * M) \times 1)$ vector. [9]

Thuật toán:

Bước 1: Cho X là một véc tơ ngẫu nhiên với các mẫu được rút ra từ các lớp c :

$$X = \{X_1, X_2, \dots, X_c\} \quad (12)$$

trong đó $X_i = \{x_1, x_2, \dots, x_n\}$

Bước 2: Các ma trận phân tán S_B và S_W được tính như sau:

$$S_B = \sum_{i=1}^c N_i (\mu_i - \mu)(\mu_i - \mu)^T \quad (13)$$

$$S_W = \sum_{i=1}^c (x_j - \mu_i)(x_j - \mu_i)^T \quad (14)$$

trong đó μ là tổng bình quân:

$$\mu = \frac{1}{N} \sum_{i=1}^N X_i \quad (15)$$

Và μ_i là bình quân của lớp $i \in \{1, \dots, c\}$:

$$\mu_i = \frac{1}{|X_i|} \sum_{x_j \in X_i} x_j \quad (16)$$

Thuật toán cổ điển Fisher hiện nay sẽ cho một phép chiếu W , nhằm tối đa hóa các tiêu chí phân chia lớp:

$$W_{opt} = \operatorname{argmax}_W \frac{|W^T S_B W|}{|W^T S_W W|} \quad (17)$$

Một giải pháp cho vấn đề tối ưu hóa này được đưa ra bằng cách giải quyết các vấn đề giá trị riêng:

$$S_B v_i = \lambda_i S_W v_i \quad (18)$$

$$S_W^{-1} S_B v_i = \lambda_i v_i \quad (19)$$

Có một vấn đề còn cần giải quyết: Việc xếp hạng của S_W là nhiều nhất $(N - c)$, với N mẫu và c lớp. Trong vấn đề nhận dạng mẫu số mẫu N là hầu như luôn luôn nhỏ hơn kích thước của dữ liệu đầu vào (số lượng thành số ít). Điều này đã được giải quyết bằng cách thực hiện một phân phân tích chính trên các dữ liệu và đưa các mẫu vào không gian $(N-c)$ chiều. Một phân tích phân biệt tuyến tính sau đó được thực hiện trên các dữ liệu giảm, bởi vì S_W không ít nữa. [9]

Các vấn đề tối ưu hóa sau đó có thể được viết lại như sau:

$$W_{pca} = \operatorname{argmax}_W |W^T S_T W| \quad (20)$$

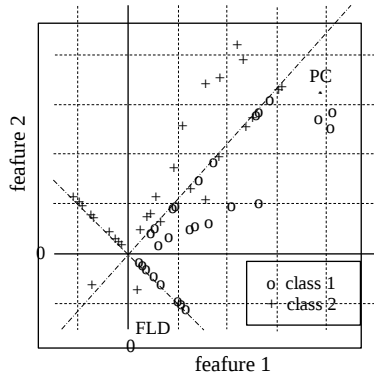
$$W_{fld} = \operatorname{argmax}_W \frac{W^T W_{pca}^T S_B W_{pca} W}{W^T W_{pca}^T S_W W_{pca} W} \quad (21)$$

Việc chuyển đổi ma trận W , mà đưa một mẫu vào không gian $(c-1)$ chiều được cho bởi:

$$W = W_{fld}^T W_{pca}^T \quad (22)$$

Fisherface tương tự Eigenface nhưng với những cải tiến trong phân loại tốt hơn về hình ảnh các lớp khác nhau. Có Fisher's Linear Discriminant (FLD), chúng ta có thể phân loại tập huấn luyện để xử lý với những người khác nhau và biểu hiện trên khuôn mặt khác nhau. Chúng ta có thể có độ chính xác tốt hơn trong biểu hiện trên khuôn mặt hơn phương pháp tiếp cận Eigenface. Bên cạnh đó, Fisherface loại bỏ ba thành phần chính đầu tiên chịu trách nhiệm những thay đổi cường độ ánh sáng, nó là bất biến hơn với cường độ ánh sáng. Fisherface phức tạp hơn Eigenface trong việc tìm kiếm các dự báo về không gian mặt. Tính tỷ lệ giữa các lớp phân tán đến phân tán

trong lớp đòi hỏi rất nhiều thời gian xử lý. Bên cạnh đó, do nhu cầu của phân loại tốt hơn, kích thước của phép chiếu trong không gian khuôn mặt là không nhỏ gọn như Eigenface, kết quả trong lưu trữ khuôn mặt lớn hơn và nhiều thời gian xử lý hơn trong nhận dạng. Phân biệt tuyến tính Fisher (FLD, Fisherface) tiếp cận bản đồ các đặc trưng không gian con hai lớp riêng biệt nhất



Hình 5. So sánh PCA và FLD trong bài toán hai lớp, nơi dữ liệu cho mỗi lớp nằm gần một không gian con tuyến tính [9]

Ưu điểm: [8]

- FLD đạt được sự phân bố giữa các lớp tốt hơn PCA
- Sử dụng ít bộ nhớ.
- Chính xác hơn phương pháp tiếp cận eigenface

B. Phân lớp biểu hiện cảm xúc khuôn mặt.

Có rất nhiều phương pháp được sử dụng để phân loại biểu cảm khuôn mặt như Mạng nơ-ron nhân tạo, “Support Vector Machine”, “Optical Flow”, “PCA”.

1) Support Vector Machines (SVM)

Xét bài toán phân lớp đơn giản nhất – phân lớp hai lớp với tập dữ liệu mẫu:

$$\{(x_i, y_i) \mid i = 1, 2, \dots, N, x_i \in \mathbb{R}^m\} \quad (23)$$

Trong đó mẫu là các vector đối tượng được phân lớp thành các mẫu dương và mẫu âm như trong hình 4:

- Các mẫu dương là các mẫu x_i thuộc lĩnh vực quan tâm và được gán nhãn $y_i = 1$.
- Các mẫu âm là các mẫu x_i không thuộc lĩnh vực quan tâm và được gán nhãn $y_i = -1$.

Thực chất phương pháp này là một bài toán tối ưu, mục tiêu là tìm ra một không gian H và siêu mặt phẳng quyết định h trên H sao cho sai số phân lớp là thấp nhất.

Trong trường hợp này, tập phân lớp SVM là *mặt siêu phẳng phân tách các mẫu dương khỏi các mẫu âm với độ chênh lệch cực đại*, trong đó *độ chênh lệch* – còn gọi là *Lề* (margin) xác định bằng khoảng cách giữa các mẫu dương và các mẫu âm

gần mặt siêu phẳng nhất. Mặt siêu phẳng này được gọi là *mặt siêu phẳng lề tối ưu*.

Các mặt siêu phẳng trong không gian đối tượng có phương trình là:

$$C + w_1 x_1 + w_2 x_2 + \dots + w_n x_n = 0 \quad (24)$$

Tương đương với công thức

$$C + \sum w_i x_i = 0 \text{ với } i=1, \dots, n \quad (25)$$

Với $w = w_1 + w_2 + \dots + w_n$ là bộ hệ số siêu phẳng hay là vector trọng số, C là độ dịch, khi thay đổi w và C thì hướng và khoảng cách từ gốc tọa độ đến mặt siêu phẳng thay đổi.

Tập phân lớp SVM được định nghĩa như sau:

$$f(x) = \text{sign}(C + \sum w_i x_i) \quad (26)$$

Trong đó

$$\text{sign}(z) = +1 \text{ nếu } z \geq 0,$$

$$\text{sign}(z) = -1 \text{ nếu } z < 0.$$

Nếu $f(x) = +1$ thì x thuộc về lớp dương (lĩnh vực được quan tâm), và ngược lại, nếu $f(x) = -1$ thì x thuộc về lớp âm (các lĩnh vực khác).

Máy học SVM là một học các siêu phẳng phụ thuộc vào tham số vector trọng số w và độ dịch C . Mục tiêu của phương pháp SVM là ước lượng w và C để cực đại hoá lề giữa các lớp dữ liệu dương và âm. Các giá trị khác nhau của lề cho ta các họ siêu mặt phẳng khác nhau, và lề càng lớn thì năng lực của máy học càng giảm. Như vậy, cực đại hoá lề thực chất là việc tìm một máy học có năng lực nhỏ nhất. Quá trình phân lớp là tối ưu khi sai số phân lớp là cực tiểu.

Ta phải giải phương trình sau:

$$\min_{w, b, \eta} C \sum_{i=1}^l \eta_i + \frac{1}{2} \|w\|^2$$

$$\text{s.t. } y_i [w \cdot x_i - b] + \eta_i \geq 1$$

$$\eta_i \geq 0, \quad i = 1, \dots, l$$

$$(27)$$

tìm ra được vector trọng số w và sai số của mỗi điểm trong tập huấn luyện là η_i từ đó ta có phương trình tổng quát của siêu phẳng tìm ra được bởi thuật toán SVM là:

$$f(x_1, x_2, \dots, x_n) = C + \sum w_i x_i \quad (28)$$

Với $i = 1, \dots, n$. Trong đó n là số dữ liệu huấn luyện. Sau khi đã tìm được phương trình của siêu phẳng bằng thuật toán SVM, áp dụng công thức này để tìm ra nhãn lớp cho các dữ liệu mới.

2) Giải thuật mạng nơ-ron lan truyền ngược

Ta sử dụng một số kí hiệu sau:

j : neuron thứ j (hay nút thứ j)

X_j : vector đầu vào của nút thứ j

W_j : vector trọng số của nút thứ j

x_{ji} : đầu vào của nút thứ j từ nút thứ i

w_{ji} : trọng số trên x_{ji}

b_j : ngưỡng tại nút thứ j

o_j : đầu ra của nút thứ j

t_j : đầu ra mong muốn của nút thứ j

Downstream(j): Tập tất cả các nút nhận đầu ra của nút thứ j làm một giá trị đầu vào.

η : tốc độ học

f : hàm truyền với $f(x) = 1 / (1 + e^{-x})$

Thuật toán lan truyền ngược được mô tả như sau:

Input:

- Mạng feed-forward với n_i đầu vào, n_h nút ẩn và n_o đầu ra.

- Hệ số học η

- Tập dữ liệu huấn luyện $D = \{ \text{là vector đầu vào, là vector đầu ra mong muốn} \}$.

Output: Các vector trọng số

Thuật toán:

Bước 1: Khởi tạo trọng số bởi các giá trị ngẫu nhiên nhỏ.

Bước 2: Lặp lại cho tới khi thỏa mãn điều kiện kết thúc.

Với mỗi mẫu, thực hiện các bước sau:

2.1 Tính đầu ra o_j cho mỗi nút j :

$$o_j = f(d - b_j) \text{ với } d = \sum x_{ji} w_{ji} \quad (29)$$

2.2 Với mỗi nút k thuộc tầng ra, tính δ_k theo công thức:

$$\delta_k = (t_k - o_k)(1 - o_k)o_k \quad (30)$$

2.3 Với mỗi nút h thuộc tầng ẩn, tính δ_h theo công thức:

$$\delta_h = o_h(1 - o_h) \sum \delta_k w_{kh} \text{ với } k \in \text{Downstream}(h) \quad (31)$$

2.4 Cập nhật: $w_{ji} = w_{ji} + \Delta w_{ji}$ (32)

Trong đó $\Delta w_{ji} = \eta \delta_k x_{ji}$

Theo hướng tiếp cận nhận dạng biểu hiện cảm xúc khuôn mặt bằng mạng nơ ron nhân tạo nhóm tác giả của bài báo [10] đã cho kết quả thực nghiệm như bảng 1.

Biểu cảm	Bình thường(50)	Vui vẻ(50)	Ngạc nhiên(50)	Buồn(50)	Trung bình
Bình thường	48	2	0	3	
Vui vẻ	1	43	0	1	
Ngạc nhiên	0	3	50	1	
Buồn	1	2	0	45	
Recall	96%	86%	100%	90%	93%
Precision	90.56%	95.6%	94.34%	93.75%	93.56%
F	0.93	0.91	0.97	0.92	0.93

Bảng 1.
Bảng thống kê kết quả thực nghiệm

Theo hướng

tiếp cận phân loại biểu hiện cảm xúc bằng Support Vector Machine nhóm tác giả của bài báo [11] đã cho kết quả thực nghiệm như bảng 2.

lab _{clac} \ lab _{ac}	an	di	fe	ha	sa	su
an	32	0	0	0	0	0
di	1	37	0	0	0	0
fe	0	0	37	0	0	0
ha	0	0	0	37	0	0
sa	4	0	0	0	37	0
su	0	0	0	0	0	37

Bảng 2. Bảng thống kê kết quả thực nghiệm

III. ĐÁNH GIÁ

AAM (active appearance model) là một thuật toán tối ưu một mô hình thống kê hình ảnh thể hiện của đối tượng vào một ảnh đầu vào mới, kết quả của quá trình tối ưu là một bộ điểm điều khiển thể hiện cấu trúc của đối tượng đã được học với các tọa độ tương ứng với thể hiện trong ảnh đầu vào của đối tượng cùng với nó là một bộ các tham số mô hình thống kê đã được ước lượng mà từ những tham số mô hình này, ta có thể dễ dàng tái cấu trúc cả về hình dạng cũng như kết cấu hình ảnh của đối tượng tương ứng một cách tương đối với thể hiện của đối tượng ở trong ảnh thử nghiệm.

Trong thực nghiệm, thuật toán AAM trích chọn các điểm điều khiển chính xác trong những trường hợp ảnh đầu vào có điều kiện đủ sáng và nền không quá phức tạp, trong trường hợp ảnh sáng không đủ tại một số vùng của khuôn mặt hay nền gần giống với màu da thì thuật toán AAM thực hiện không hiệu quả

TÀI LIỆU THAM KHẢO

- [1] Vural, E., et al., "Automated Drowsiness Detection For Improved Driving safety", in Proc. Of the Int.Conf. on Automotive Technologies, 2008.
- [2] Ashraf, A.B., et al., "The painful face-Pain expression recognition using active appearance models", Image and Vision Computing, 2009. 27(12): pp. 1788-1796
- [3] P. Ekman, D. Matsumoto, W.V. Friesen, "Facial Expression in Affective Disorders, What the Face Reveals", P. Ekman and E.L. Rosenberg, eds., pp.429-439, 2005.
- [4] Lê Thị Kim Nga, Phạm Trần Thiện, Hà Mạnh Toàn, Lâm Thành Hiền, "Một kỹ thuật định vị các điểm điều khiển trên khuôn mặt dựa trên mô hình chất liệu", Kỷ yếu Hội nghị Quốc gia lần thứ VIII về Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin (FAIR), Hà Nội, 2015

- [5] Garcia, C. and G. Tziritas. "Face detection using quantized skin color regions merging and wavelet packet analysis." IEEE Trans. Multimedia, 1999. pp. 264 - 277.
- [6] Lê Hoàng Thái, Lê Minh Trí, Nguyễn Thủy Hằng (2007) “Đề xuất phương pháp trích chọn đặc trưng mới cho bài toán nhận dạng mặt người”, Kỷ yếu Hội thảo Quốc gia “Một số vấn đề chọn lọc của Công nghệ thông tin và Truyền thông”, Đà Lạt 15-17/6/2006, tr. 2-5.
- [7] M. A. Turk và A. P. Pentland - Face Recognition Using Eigenfaces, Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 586-591, June 1991.
- [8] Mạch Thị Kim Hạnh, “Xác thực sinh trắc học khuôn mặt trên thiết bị di động”, Luận văn Thạc sỹ Công nghệ thông tin, trường Đại học Lạc Hồng, 2013.
- [9] Fisherface, http://docs.opencv.org/modules/contrib/doc/facerec/facerec_tutorial.html
- [10] Lê Thái Linh, Nguyễn Linh Giang, Phan Đăng Hưng, “Nhận dạng biểu cảm gương mặt sử dụng mạng nơ-ron và đường cong bezier”, Kỷ yếu Hội nghị Quốc gia lần thứ VIII về Nghiên cứu cơ bản và ứng dụng Công Nghệ thông tin (FAIR), 2013
- [11] Kotsia, I., & Pitas, I. (2005, September). Real time facial expression recognition from image sequences using support vector machines. In Image Processing, 2005. ICIP 2005. IEEE International Conference on (Vol. 2, pp. II-966). IEEE.

Abstract: Expression recognition (happy, sad, disgust, surprise, angry, fear expressions) is application of advanced object detection. This has found applications in technical fields such as Human-computer-Interaction (HCI) and security monitoring. It generally requires fast processing and decision making. This paper was presented a method for face detection based on material models and emotion recognition.



Th.S Huỳnh Cao Tuấn ,
 Nhận học vị Thạc sỹ năm 2012. Hiện công tác tại Trường Đại học Lạc Hồng.
 Lĩnh vực nghiên cứu: Công nghệ robot, khai phá dữ liệu, xử lý ảnh, học máy.



TS Nguyễn Thanh Bình ,
 Nhận học vị Tiến sỹ năm 1995. Hiện công tác tại Học viện Bưu chính Viễn Thông.
 Lĩnh vực nghiên cứu: Kỹ thuật phát thanh và truyền hình, xử lý âm thanh, xử lý hình ảnh.



PGS.TS Đỗ Năng Toàn,
 Nhận học vị Tiến sỹ năm 2001, học hàm Phó giáo sư năm 2007. Hiện công tác tại Viện Công nghệ Thông tin – Đại học Quốc gia Hà Nội. Lĩnh vực nghiên cứu: Trí tuệ nhân tạo, xử lý ảnh, học máy