

## MỤC LỤC

MỤC LỤC.....	1
DANH MỤC HÌNH VẼ.....	3
LỜI CẢM ƠN.....	5
LỜI MỞ ĐẦU.....	6
CHƯƠNG 1. KHÁI QUÁT VỀ XỬ LÝ ẢNH VÀ BÀI TOÁN PHÁT HIỆN MẶT NGƯỜI.....	9
1.1. Khái quát về xử lý ảnh.....	9
1.1.1. Một số khái niệm cơ bản.....	9
1.1.2. Một số vấn đề trong xử lý ảnh.....	10
1.1.3. Một số ứng dụng trong xử lý ảnh.....	12
1.2. Bài toán phát hiện mặt người.....	13
1.2.1. Giới thiệu về bài toán xác định mặt người trong ảnh.....	13
1.2.2. Định nghĩa bài toán xác định mặt người.....	13
CHƯƠNG 2. BÀI TOÁN NHẬN DIỆN MẶT NGƯỜI.....	14
2.1. Bài toán nhận dạng mặt người.....	14
2.2. Những khó khăn của nhận dạng khuôn mặt.....	14
2.3. Tầm quan trọng của bài toán nhận diện mặt người.....	15
2.4. Các ứng dụng đặc trưng của bài toán nhận diện mặt người.....	16
Các ứng dụng khác:.....	17
2.5. Xây dựng hệ thống nhận diện mặt người đặc trưng.....	17
2.6. Một số phương pháp nhận diện mặt người.....	18
2.6.1. Dựa trên tri thức.....	19
2.6.2. Hướng tiếp cận dựa trên đặc trưng không thay đổi.....	22
2.6.3. Hướng tiếp cận dựa trên so khớp mẫu.....	25
2.6.4. Hướng tiếp cận dựa trên diện mạo.....	28
CHƯƠNG 3. CƠ SỞ LÝ THUYẾT.....	35
3.1. Phát hiện mặt người.....	35

3.1.1. Tiếp cận Boosting.....	35
3.1.2. Adaboost.....	36
3.1.3. Các đặc trưng Haar-Like.....	40
3.1.4. Cascade of Classifiers.....	43
3.1.5. Cascade of boosting classifiers.....	45
3.2. Nhận diện mặt người.....	46
3.2.1. Nhận dạng dựa trên mối quan hệ giữa các phần tử (Feature Based).....	46
3.2.2. Nhận dạng dựa trên xét toàn diện khuôn mặt (Appearance Based).....	47
3.2.3. Nội dung phương pháp PCA.....	49
3.3. Sơ đồ khối của hệ thống.....	52
3.3.1. Sơ đồ khối phần cứng.....	52
3.3.2. Sơ đồ khối phần mềm.....	53
3.3.3. Thiết kế hệ thống cho chương trình nhận diện khuôn mặt.....	53
CHƯƠNG 4. XÂY DỰNG THỰC NGHIỆM.....	55
4.1. Sơ lược về OpenCV.....	55
4.2. Sơ lược về phần cứng Kit Raspberry Pi 3.....	56
4.3. Sơ lược về bộ điều khiển mở cửa Rơ-le.....	57
4.4. Sơ lược Webcam Logitech C270.....	59
4.5. Kết quả thực nghiệm.....	60
CHƯƠNG 5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN.....	62
5.1. Kết luận.....	62
5.2. Hướng phát triển.....	62
TÀI LIỆU THAM KHẢO.....	64

## **DANH MỤC HÌNH VẼ**

Hình 1. Quá trình xử lý ảnh

Hình 2. Các bước cơ bản trong một hệ thống xử lý ảnh

Hình 3. Hệ thống nhận diện mặt người đặc trưng

Hình 4: Một phương pháp xác định khuôn mặt diễn theo hướng tiếp cận top-down

Hình 5: Phương pháp chiếu

Hình 6. Kết cấu khuôn mặt

Hình 7. Màu sắc da mặt

Hình 8: Một mẫu khuôn mặt

Hình 9. Các vector quan sát để huấn luyện cho HMM

Hình 10. Các trạng thái ẩn

Hình 11. Xác định khuôn mặt bằng HMM

Hình 12. Boosting

Hình 13. Sơ đồ khối thuật toán AdaBoost

Hình 14. Đặt trung Haar-like cơ bản

Hình 15. Các đặc trưng mở rộng của các đặc trưng Haar-like cơ sở

Hình 16. Cách tính Integral Image của ảnh

Hình 17. Ví dụ cách tính nhanh tổng các điểm ảnh của vùng D trên ảnh

Hình 18. Cascade of classifiers

Hình 19. Cascade of boosting classifiers

Hình 20. Hệ thống phát hiện mặt người trong ảnh sử dụng Cascade of classifier

Hình 21. Lưu đồ thuật toán tính đặc trưng mặt trong tập huấn luyện

Hình 22. Lưu đồ thuật toán tính nhận diện hình ảnh khuôn mặt

Hình 23. Sơ đồ khối cho phần cứng

Hình 24. Sơ đồ khối cho ứng dụng

Hình 25. Các thành phần chương trình nhận diện mặt người

Hình 26. Cấu trúc cơ bản của OpenCV

Hình 27. Raspberry Pi 3

Hình 28. Bộ điều khiển Rơ-le

Hình 29. Webcam Logitech C270

Hình 30. Nhận diện là người thân

Hình 31. Nhận diện không phải là người thân

## LỜI CẢM ƠN

Để hoàn thành đồ án này, em xin tỏ lòng biết ơn sâu sắc đến Cô ThS. Hồ Thị Kim Hoàng, đã tận tình hướng dẫn trong suốt quá trình viết Báo cáo đồ án. Em chân thành cảm ơn quý Thầy, Cô trong khoa Kỹ Thuật Máy Tính, Trường Đại Học Công Nghệ Thông Tin TP. Hồ Chí Minh đã tận tình truyền đạt kiến thức trong những năm em học tập.

Với vốn kiến thức được tiếp thu trong quá trình học không chỉ là nền tảng cho quá trình nghiên cứu đồ án mà còn là hành trang quý báu để em bước vào đời một cách vững chắc và tự tin. Và đặc biệt, trong học kỳ này. Nếu không có những lời hướng dẫn, dạy bảo của các thầy cô thì em nghĩ đồ án này của em rất khó có thể hoàn thiện được.

Một lần nữa, em xin chân thành cảm ơn cô. Bài báo cáo đồ án thực hiện trong khoảng thời gian gần hơn hai tháng. Bước đầu tìm hiểu của chúng em còn hạn chế và còn nhiều bỏ ngỏ.

Đồng thời do trình độ cũng như kinh nghiệm thực tiễn còn hạn chế nên bài báo cáo không thể tránh khỏi những thiếu sót, em rất mong nhận được ý kiến đóng góp của Cô để em học thêm được nhiều kinh nghiệm và sẽ hoàn thành tốt hơn bài luận văn tốt nghiệp sắp tới.

Em xin chân thành cảm ơn!

## LỜI MỞ ĐẦU

Hiện nay, vấn đề kiểm soát an ninh đang là một trong những vấn đề được quan tâm hàng đầu trên toàn thế giới. Bài toán nhận dạng vì vậy được quan tâm rất nhiều. Trong vấn đề nhận dạng con người, đã có nhiều phương pháp được đưa ra và có độ chính xác cao. Các bài toán nhận dạng được áp dụng theo hai con đường: nhận dạng offline và nhận dạng online.

- Nhận dạng offline: đối tượng được thu thập thành các file dữ liệu và được chuyển về trung tâm. Tại đó, các số liệu sẽ được phân tích xử lý, thời xử lý không yêu cầu quá khắt khe.
- Nhận dạng online: đối tượng được thu thập và xử lý trực tiếp tại hiện trường. Thời gian xử lý luôn được yêu cầu khắt khe để có thể bám theo được đối tượng.

Một bài toán nhận dạng con người đang phổ biến hiện nay:

- Nhận dạng vân tay
- Nhận dạng bàn tay
- Nhận dạng vân mắt
- Sử dụng thẻ xác nhận ...

Các phương án nhận dạng trên đã qua một thời gian dài phát triển nên đã đạt được độ chính xác cao. Tuy nhiên, các phương pháp này vẫn còn gặp phải một số hạn chế như sau:

- Các hệ thống trên đòi hỏi người sử dụng phải tiếp xúc trực tiếp với hệ thống, người đó ý thức được rằng mình đang được nhận dạng.
- Với các hệ thống thẻ xác nhận, người nhận biết yêu cầu phải có thẻ, điều này chỉ có thể áp dụng với các thành viên thường xuyên ra vào hệ thống.
- Với các hệ thống vân mắt, vân tay,... không phải bao giờ ta cũng có thể lấy được mẫu của đối tượng.
- Nhận dạng dựa vào vân tay, bàn tay, vân mắt, thẻ xác nhận,... không phải là cách tự nhiên mà con người sử dụng để nhận biết...

Hiện nay, với sự phát triển của toán học và khoa học máy tính, một trong các bài toán nhận dạng đang được quan tâm nghiên cứu nhiều là nhận dạng khuôn mặt. Mặc dù

nhận dạng khuôn mặt hiện nay chưa đạt được độ chính xác cao như một số phương pháp khác đã nêu trên (ví dụ nhận dạng vân tay), nhưng nhận dạng khuôn mặt vẫn luôn nhận được nhiều sự quan tâm của khoa học vì một số các lý do như sau:

- Nó gắn liền với cách mà con người nhận biết.
- Với các hệ thống camera số, ta có thể dễ dàng lấy được thông tin về đối tượng mà không cần tiếp xúc trực tiếp.
- Giám sát được đối tượng một cách kín đáo.
- Các thuật toán phức tạp được hỗ trợ bởi các hệ máy tính có tốc độ cao...

Nhận dạng khuôn mặt offline đã có nhiều công trình công bố và các thuật toán có độ chính xác tương đối cao. Tuy nhiên để kết hợp với các hệ thống khác để tạo ra một hệ thống nhận dạng online thì vẫn chưa đạt được kết quả mong đợi. Hệ thống nhận dạng mặt người online sẽ mang lại thêm một giải pháp kiểm soát an ninh phù hợp với các điều kiện mà các hệ thống nhận dạng khác không thỏa mãn được. Với hệ thống nhận dạng mặt người online có được độ tin cậy cao hơn, thì các điều kiện đầu vào nên được quy định chặt chẽ hơn.

- Đồ án được thực hiện với mục đích nghiên cứu xây dựng một hệ thống nhận dạng mặt người online như trên.

Yêu cầu đặt ra của đồ án như sau: xây dựng một hệ thống nhận dạng mặt người online với đầu vào là ảnh khuôn mặt người cần nhận dạng, được thu thập bằng camera số, đầu ra là kết quả nhận dạng cho khuôn mặt đó.

- Nhiệm vụ đặt ra của đồ án là giải quyết yêu cầu trên.

Để có thể xử lý được bài toán trên, đồ án đề xuất hướng tiếp cận là sử dụng các kỹ thuật xử lý ảnh số. Các ảnh số được thu thập bằng camera số và xử lý thông qua máy tính PC. Đây là hệ thống nhận dạng online nên có yêu cầu về thời gian.

Bài toán nhận dạng mặt người online là sự kết hợp của 2 bài toán:

- Bài toán xác định mặt người trong ảnh số (Human Face Detection)
- Bài toán nhận dạng mặt người (Face Recognition)

Trong hai bài toán này, bài toán xác định mặt người là bài toán mang tính quyết định tới độ chính xác của hệ thống. Khuôn mặt được xác định “tốt” sẽ mang lại kết quả cao cho quá trình nhận dạng. Việc lựa chọn phương án giải quyết hai bài toán trên phải tính toán tới thời gian xử lý. Do hệ thống là online, nên để có thể theo dõi được đối tượng thì thời gian xử lý phải đủ nhanh.

Với mục tiêu chính là tìm hiểu thông tin, các bài toán và giải pháp về nhận diện khuôn mặt, đồng thời áp dụng vào bài toán thiết kế “Robot nhận diện người thân”, đồ án được trình với những nội dung dưới đây.

## **CHƯƠNG 1. KHÁI QUÁT VỀ XỬ LÝ ẢNH VÀ BÀI TOÁN PHÁT HIỆN MẶT NGƯỜI**

### **1.1. Khái quát về xử lý ảnh**

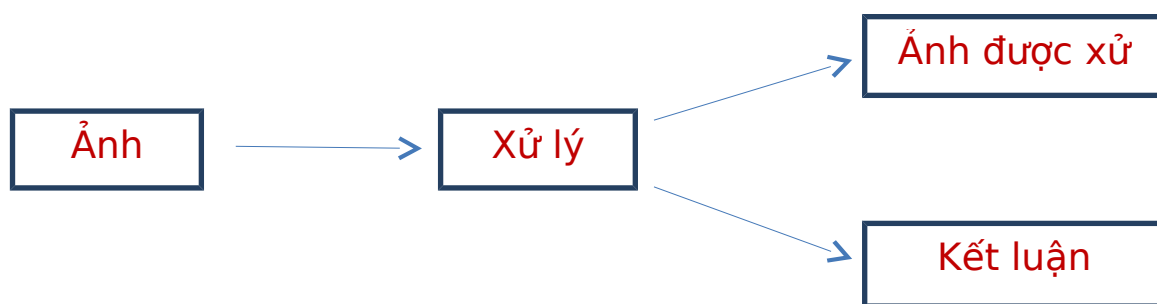
#### **1.1.1. Một số khái niệm cơ bản**

Xử lý ảnh là một trong những mảng quan trọng nhất trong kỹ thuật thị giác máy tính, là tiền đề cho nhiều nghiên cứu thuộc lĩnh vực này. Hai nhiệm vụ cơ bản của quá trình xử lý ảnh là nâng cao chất lượng thông tin hình ảnh và xử lý số liệu cung cấp cho các quá trình khác trong đó có việc ứng dụng thị giác vào điều khiển.



Quá trình bắt đầu từ việc thu nhận ảnh nguồn (từ các thiết bị thu nhận ảnh dạng số hoặc tương tự) gửi đến máy tính. Dữ liệu ảnh được lưu trữ ở định dạng phù hợp với quá trình xử lý. Người lập trình sẽ tác động các thuật toán tương ứng lên dữ liệu ảnh nhằm thay đổi cấu trúc ảnh phù hợp với các ứng dụng khác nhau.

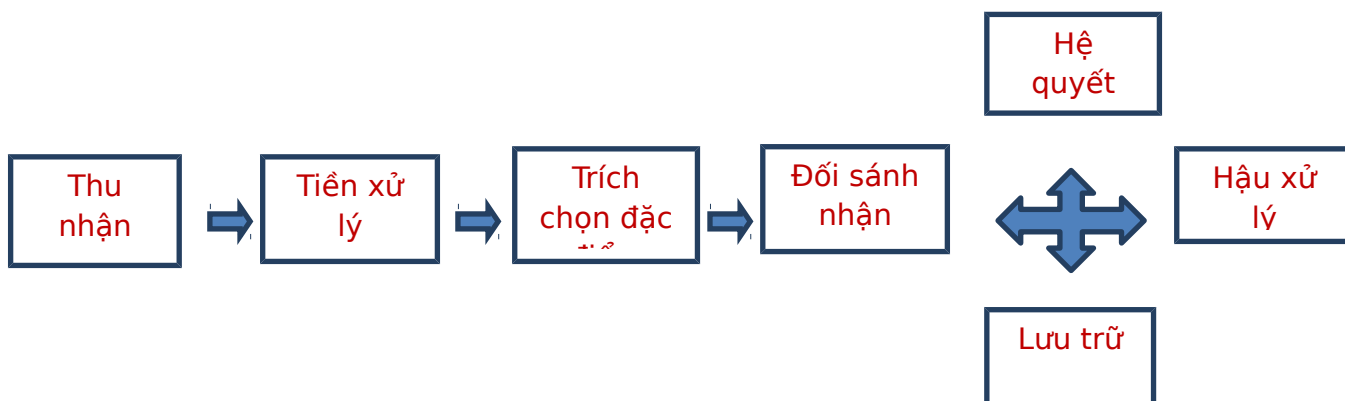
Quá trình xử lý nhận dạng ảnh được xem như là quá trình thao tác ảnh đầu vào nhằm cho ra kết quả mong muốn. Kết quả đầu ra của một quá trình xử lý ảnh có thể là một ảnh “đã được xử lý” hoặc một kết luận.



Hình 1. Quá trình xử lý ảnh

Ảnh có thể xem là tập hợp các điểm ảnh và mỗi điểm ảnh được xem như là đặc trưng cường độ sáng hay một dấu hiệu nào đó tại một vị trí nào đó của đối tượng trong không gian và nó có thể xem như một hàm  $n$  biến  $P(c, c_1, c_2, \dots)$ . Do đó, ảnh trong xử lý ảnh có thể xem như ảnh  $n$  chiều.

\* Sơ đồ tổng quát của một hệ thống xử lý ảnh:



Hình 2. Các bước cơ bản trong một hệ thống xử lý ảnh

### 1.1.2. Một số vấn đề trong xử lý ảnh

#### 1.1.2.1. Các hệ thống xử lý ảnh

##### a. Tiền xử lý:

Tiền xử lý là giai đoạn đầu tiên trong xử lý ảnh số. Tùy thuộc vào quá trình xử lý tiếp theo trong giai đoạn này sẽ thực hiện các công đoạn khác nhau như: nâng cấp, khôi phục ảnh, nắn chỉnh hình học, khử nhiễu v.v..

##### b. Trích chọn đặc điểm:

Các đặc điểm của đối tượng được trích chọn tùy theo mục đích nhận dạng trong quá trình xử lý ảnh. Trích chọn hiệu quả các đặc điểm giúp cho việc nhận dạng các đối tượng ảnh chính xác, với tốc độ tính toán cao và dung lượng nhớ lưu trữ giảm.

##### c. Đối sánh, nhận dạng:

Nhận dạng tự động (automatic recognition), mô tả đối tượng, phân loại và phân nhóm các mẫu là những vấn đề quan trọng trong thị giác máy, được ứng dụng trong nhiều ngành khoa học khác nhau. Ví dụ mẫu có thể là ảnh của vân tay, ảnh của một vật nào đó được chụp, một chữ viết, khuôn mặt người hoặc một ký đồ tín hiệu tiếng nói. Khi biết một mẫu nào đó, để nhận dạng hoặc phân loại mẫu đó. Hệ thống nhận dạng tự động bao gồm ba khâu tương ứng với ba giai đoạn chủ yếu sau đây:

- Thu nhận dữ liệu và tiền xử lý.
- Biểu diễn dữ liệu.
- Nhận dạng, ra quyết định.

Bốn cách tiếp cận khác nhau trong lý thuyết nhận dạng là:

- Đối sánh mẫu dựa trên các đặc trưng được trích chọn.
- Phân loại thống kê.
- Đối sánh cấu trúc.
- Phân loại dựa trên mạng nơ-ron nhân tạo.

Trong các ứng dụng rõ ràng là không thể chỉ dùng có một cách tiếp cận đơn lẻ để phân loại “tối ưu” do vậy cần sử dụng cùng một lúc nhiều phương pháp và cách tiếp cận khác nhau. Do vậy, các phương thức phân loại tổ hợp hay được sử dụng khi nhận dạng và nay đã có những kết quả có triển vọng dựa trên thiết kế các hệ thống lai (hybrid system) bao gồm nhiều mô hình kết hợp.

### **1.1.2.2. Các hình thái của ảnh**

#### **a.Chuyển ảnh màu thành ảnh xám:**

Đơn vị tế bào của ảnh số là pixel. Tùy theo mỗi định dạng là ảnh màu hay ảnh xám mà từng pixel có thông số khác nhau. Đối với ảnh màu từng pixel sẽ mang thông tin của ba màu cơ bản tạo ra bản màu khả kiến là Đỏ (R), Xanh lá (G) và Xanh biển (B) [Thomas 1892]. Trong mỗi pixel của ảnh màu, ba màu cơ bản R, G và B được bố trí sát nhau và có cường độ sáng khác nhau. Thông thường, mỗi màu cơ bản được biểu diễn bằng tám bit tương ứng 256 mức độ màu khác nhau. Như vậy mỗi pixel chúng ta sẽ có  $28 \times 3 = 224$  màu (khoảng 16.78 triệu màu). Đối với ảnh xám, thông thường mỗi pixel mang thông tin của 256 mức xám (tương ứng với tám bit) như vậy ảnh xám hoàn toàn có thể tái hiện đầy đủ cấu trúc của một ảnh màu tương ứng thông qua tám mặt phẳng bit theo độ xám.

#### **b.Lược đồ xám của ảnh (Histogram):**

Lược đồ xám của một ảnh số có các mức xám trong khoảng  $[0, L-1]$  là một hàm rời rạc  $p(r_k) = n_k/n$ . Trong đó  $n_k$  là số pixel có mức xám thứ  $r_k$ ,  $n$  là tổng số pixel trong ảnh và  $k=0,1,2,\dots,L-1$ . Do đó  $P(r_k)$  cho một xấp xỉ xác suất xảy ra mức xám  $r_k$ . Vẽ hàm này với tất cả các giá trị của  $k$  sẽ biểu diễn khái quát sự xuất hiện các mức xám của một ảnh. Chúng ta cũng có thể thể hiện lược đồ mức xám của ảnh thông qua tần suất xuất hiện mỗi mức xám qua hệ tọa độ vuông góc xOy. Trong đó, trục hoành biểu diễn số mức xám từ 0 đến N (số bit của ảnh xám). Trục tung biểu diễn số pixel của mỗi mức xám.

### **1.1.3. Một số ứng dụng trong xử lý ảnh**

Như đã nói ở trên, các kỹ thuật xử lý ảnh trước đây chủ yếu được sử dụng để nâng cao chất lượng hình ảnh, chính xác hơn là tạo cảm giác về sự gia tăng chất lượng

ảnh quang học trong mắt người quan sát. Thời gian gần đây, phạm vi ứng dụng xử lý ảnh mở rộng không ngừng, có thể nói hiện không có lĩnh vực khoa học nào không sử dụng các thành tựu của công nghệ xử lý ảnh số.

Trong y học các thuật toán xử lý ảnh cho phép biến đổi hình ảnh được tạo ra từ nguồn bức xạ X-ray hay nguồn bức xạ siêu âm thành hình ảnh quang học trên bề mặt film x-quang hoặc trực tiếp trên bề mặt màn hình hiển thị. Hình ảnh các cơ quan chức năng của con người sau đó có thể được xử lý tiếp để nâng cao độ tương phản, lọc, tách các thành phần cần thiết (chụp cắt lớp) hoặc tạo ra hình ảnh trong không gian ba chiều (siêu âm 3 chiều).

Trong lĩnh vực địa chất, hình ảnh nhận được từ vệ tinh có thể được phân tích để xác định cấu trúc bề mặt trái đất. Kỹ thuật làm nổi đường biên (image enhancement) và khôi phục hình ảnh (image restoration) cho phép nâng cao chất lượng ảnh vệ tinh và tạo ra các bản đồ địa hình 3-D với độ chính xác cao.

Trong ngành khí tượng học, ảnh nhận được từ hệ thống vệ tinh theo dõi thời tiết cũng được xử lý, nâng cao chất lượng và ghép hình để tạo ra ảnh bề mặt trái đất trên một vùng rộng lớn, qua đó có thể thực hiện việc dự báo thời tiết một cách chính xác hơn.

Xử lý ảnh còn được sử dụng rộng rãi trong lĩnh vực hình sự và các hệ thống bảo mật hoặc kiểm soát truy cập: quá trình xử lý ảnh với mục đích nhận dạng vân tay hay khuôn mặt cho phép phát hiện nhanh các đối tượng nghi vấn cũng như nâng cao hiệu quả hệ thống bảo mật cá nhân cũng như kiểm soát ra vào. Ngoài ra, có thể kể đến các ứng dụng quan trọng khác của kỹ thuật xử lý ảnh tĩnh cũng như ảnh động trong đời sống như tự động nhận dạng, nhận dạng mục tiêu quân sự, máy nhìn công nghiệp trong các hệ thống điều khiển tự động, nén ảnh tĩnh, ảnh động để lưu và truyền trong mạng viễn thông v.v.

### 1.2. Bài toán phát hiện mặt người

#### 1.2.1. Giới thiệu về bài toán xác định mặt người trong ảnh.

Trong nhiều năm qua, có rất nhiều công trình nghiên cứu về bài toán nhận dạng mặt người. Các nghiên cứu đi từ bài toán đơn giản, từ việc nhận dạng một mặt người trong ảnh đen trắng cho đến mở rộng cho ảnh màu và có nhiều mặt người trong ảnh. Đến nay các bài toán xác định mặt người đã mở rộng với nhiều miền nghiên cứu như nhận dạng khuôn mặt, định vị khuôn mặt, theo dõi mặt người hay nhận dạng cảm xúc mặt người...

Phát hiện mặt người trong ảnh là phần đầu tiên của một hệ thống nhận dạng mặt người. Các hệ thống nhận dạng khuôn mặt được bắt đầu xây dựng từ những năm 1970, tuy nhiên do còn hạn chế về các luật xác định mặt người nên chỉ được áp dụng trong một số ứng dụng như nhận dạng thẻ căn cước. Nó chỉ được phát triển mạnh mẽ từ những năm 1990 khi có những tiến bộ trong công nghệ video và ngày nay thì các ứng dụng của xác định mặt người đã trở nên phổ biến trong cuộc sống.

#### 1.2.2. Định nghĩa bài toán xác định mặt người.

Xác định khuôn mặt người là một kỹ thuật máy tính để xác định các vị trí và kích thước của các khuôn mặt người trong các ảnh bất kì. Kỹ thuật này nhận biết các đặc trưng của khuôn mặt và bỏ qua những thứ khác như: tòa nhà, cây cối, cơ thể ...

## CHƯƠNG 2. BÀI TOÁN NHẬN DIỆN MẶT NGƯỜI

### 2.1. Bài toán nhận dạng mặt người

Hệ thống nhận dạng mặt người là một hệ thống nhận vào là một ảnh hoặc một đoạn video (một chuỗi các ảnh). Qua xử lý tính toán hệ thống xác định được vị trí mặt người trong ảnh (nếu có) và xác định là người nào trong số những người hệ thống đã được biết (qua quá trình học) hoặc là người lạ.

### 2.2. Những khó khăn của nhận dạng khuôn mặt

a. Tư thế góc chụp: Ảnh chụp khuôn mặt có thể thay đổi rất nhiều bởi vì góc chụp giữa camera và khuôn mặt. Chẳng hạn như : chụp thẳng, chụp xéo bên trái 45 độ hay xéo bên phải 45 độ, chụp từ trên xuống, chụp từ dưới lên, v.v...). Với các tư thế khác nhau, các thành phần trên khuôn mặt như mắt, mũi, miệng có thể bị khuất một phần hoặc thậm chí khuất hết. Khuôn mặt đang nhìn thẳng nhưng góc chụp của máy ảnh lại lệch nhiều so với hướng nhìn thẳng của khuôn mặt hoặc là lúc chụp ảnh mặt người quay nghiêng sang một bên nào đó nhiều đều là những khó khăn rất lớn trong bài toán nhận diện mặt người.

b. Sự xuất hiện hoặc thiếu một số thành phần của khuôn mặt: Các đặc trưng như râu mép, râu hàm, mắt kính, v.v... có thể xuất hiện hoặc không. Vấn đề này làm cho bài toán càng khó khăn hơn rất nhiều.

c. Sự biểu cảm của khuôn mặt: Biểu cảm của khuôn mặt người có thể làm ảnh hưởng đáng kể lên các thông số của khuôn mặt. Chẳng hạn, cùng một khuôn mặt một người, nhưng có thể sẽ rất khác khi họ cười hoặc sợ hãi....

d. Sự che khuất: Khuôn mặt có thể bị che khuất bởi các đối tượng khác hoặc các khuôn mặt khác.

e. Hướng của ảnh: Các ảnh của khuôn mặt có thể biến đổi rất nhiều với các góc quay khác nhau của trục camera. Chẳng hạn chụp với trục máy ảnh nghiêng làm cho khuôn mặt bị nghiêng so với trục của ảnh.

f. Điều kiện của ảnh: Ảnh được chụp trong các điều kiện khác nhau về: chiếu sáng, về tính chất camera (máy kỹ thuật số, máy hồng ngoại, v.v...) ảnh hưởng rất nhiều đến chất lượng ảnh khuôn mặt.

g. Nền ảnh phức tạp: nền của ảnh phức tạp là một trong những khó khăn nhất trong bài toán nhận diện khuôn mặt người trong ảnh, khuôn mặt người sẽ dễ bị nhầm lẫn với nhiều khung cảnh phức tạp xung quanh và ảnh hưởng rất nhiều đến quá trình phân tích và rút trích các đặc trưng của khuôn mặt trong ảnh, có thể dẫn đến không

nhận ra khuôn mặt hoặc là nhận nhầm các khung cảnh xung quanh thành khuôn mặt người.

h. Màu sắc của da mặt: màu sắc của da mặt quá tối hoặc gần với màu sắc của khung cảnh môi trường xung quanh cũng là một khó khăn với bài toán nhận diện mặt người. Nếu màu sắc của da người quá tối thì thuật toán sẽ gặp khó khăn trong việc nhận diện các đặc trưng và có thể không tìm ra được khuôn mặt người.

### **2.3. Tầm quan trọng của bài toán nhận diện mặt người**

Nếu như công nghệ nhận dạng giọng nói phù hợp với các ứng dụng callcenter và nhược điểm của nó là tiếng ồn, không phù hợp với nơi công cộng đông người thì công nghệ nhận dạng chữ ký cũng gây nhiều phiền phức cho người sử dụng vì khó duy trì được chữ ký giống nhau ngay trong cùng một thời điểm. Trong khi đó, công nghệ nhận dạng vân tay hiện đã được dùng khá phổ biến, nhưng cũng có nhược điểm là bị ảnh hưởng bởi độ ẩm da và đặc biệt sẽ không chính xác cao với người có tay hay tiếp xúc với hoá chất. Bên cạnh đó, nhận dạng bằng vân tay còn có những hạn chế về việc giả mạo, thậm chí kẻ gian có thể sử dụng tay của người khác để xác thực. Còn công nghệ nhận dạng mống mắt là một công nghệ có nhiều ưu điểm về độ chính xác cũng như khả năng chống giả mạo. Việc xác thực, nhận dạng trong công nghệ này dựa trên sơ đồ mạch máu trong võng mạc mắt. Sơ đồ này rất ít thay đổi kể từ lúc mới sinh cho tới khi già và đặc biệt là khi chết, hệ thống sơ đồ mạch máu này sẽ biến mất. Vì thế, không thể dùng người chết để xác thực cho việc truy cập trái phép. Tuy nhiên, công nghệ này lại bị tác động bởi nhiều yếu tố khác như độ rộng của mắt, lông mi, kính đeo và khó triển khai phổ biến trên diện rộng do độ phức tạp của các thiết bị.

Trong khi đó, mặc dù độ chính xác không phải ưu điểm nổi bật của công nghệ nhận dạng mặt người song không giống với những phương pháp khác, nhận dạng mặt người là công nghệ nhận dạng không cần đến sự tiếp xúc trực tiếp giữa đối tượng và thiết bị thu nhận. Thay vào đó, công nghệ này mang tính theo dõi, giám sát và rất thuận tiện cho những ứng dụng bảo vệ mục tiêu, chống khủng bố tại những điểm công cộng đông người. Đây cũng là ưu điểm nổi trội của nhận dạng mặt người mà các công nghệ nhận dạng khác khó có thể có được.

### 2.4. Các ứng dụng đặc trưng của bài toán nhận diện mặt người

Bài toán phát hiện mặt người trong ảnh là một bài toán phức tạp nhưng lại hấp dẫn rất nhiều người nghiên cứu về nó, có lẽ bởi tính ứng dụng thực tế của bài toán là rất lớn và đa dạng trong cuộc sống hiện đại. có thể kể ra một số ứng dụng thông dụng nhất của bài toán nhận diện mặt người:

- *Hệ thống giao tiếp thông minh giữa người và máy*: con người có thể xây dựng những hệ thống thông giao tiếp giữa người và máy tính thông qua việc nhận diện khuôn mặt, biểu cảm trên khuôn mặt người để dự đoán, nhận biết trạng thái tâm lí hiện thời của người đó. Một ngôi nhà thông minh trong tương lai có thể nhận biết được chủ nhân của nó thông qua nhận biết khuôn mặt, dáng người, giao tiếp qua giọng nói, vân tay...
- *Nhận dạng tội phạm*: hệ thống có thể nhận diện ra một khuôn mặt ngay tức thì và đối chiếu với hàng triệu bản ghi có sẵn trong cơ sở dữ liệu để chỉ ra đó có thể là một tội phạm đang truy tìm hay không, hoặc đó có thể là một nhân vật nào đó đặc biệt cần quan tâm..vv..
- *Giải trí*: trong hầu hết các máy ảnh hiện đại ngày nay đều có chức năng tự động nhận diện mặt người để có thể lấy độ nét, điều chỉnh ánh sáng cho phù hợp với khung cảnh xung quanh. Trên một số trang web cũng đã áp dụng công nghệ tự động nhận diện mặt người và so sánh với kho dữ liệu khổng lồ của mình để đưa ra những lời chào, dịch vụ thông minh nhất cho người sử dụng.
- *Hệ thống quan sát, theo dõi và bảo vệ*: các hệ thống camera sẽ xác định đâu là con người và theo dõi người đó...
- *Điều khiển ra vào các cơ quan, văn phòng*: cho phép nhân viên ra vào các khu vực quan trọng mà không cần phải đăng nhập hay dùng thẻ. Nếu kết hợp với sử dụng vân tay hay hốc mắt thì sẽ đem lại kết quả chính xác cao.
- *Tổ chức tìm kiếm liên quan đến con người*: thông qua khuôn mặt trên nhiều hệ cơ sở dữ liệu lớn.
- *Phân tích cảm xúc*: khuôn mặt người..v..v..

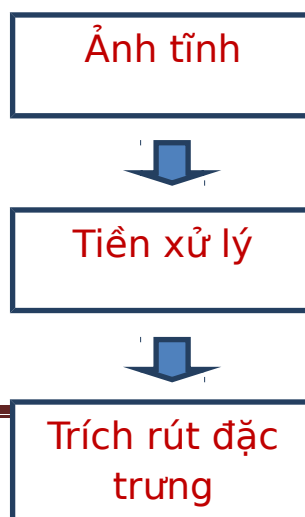


### Các ứng dụng khác:

- An ninh sân bay, xuất nhập cảnh (hiện nay cơ quan xuất nhập cảnh Mỹ đã áp dụng ). Dùng camera quan sát để xác thực người nhập cảnh và kiểm tra xem người đấy có phải là tội phạm hay phân tử khủng bố không.
- Tìm kiếm và tổ chức dữ liệu liên quan đến con người thông qua khuôn mặt người trên nhiều hệ cơ sở dữ liệu lưu trữ thật lớn, như internet, các hãng truyền hình, .... Ví dụ: tìm các đoạn video có tổng thống Bush phát biểu, tìm các phim có diễn viên Lý Liên Kiệt đóng, tìm các trận đá bóng có Ronaldo đá, ...
- Kiểm tra trạng thái người lái xe có ngủ gật, mất tập trung hay không, và hỗ trợ thông báo khi cần thiết.
- Tương lai sẽ phát triển các loại thẻ thông minh có tích hợp sẵn đặc trưng của người dùng trên đó, khi bất cứ người dùng khác dùng để truy cập hay xử lý tại các hệ thống sẽ được yêu cầu kiểm tra các đặc trưng khuôn mặt so với thẻ để biết nay có phải là chủ thẻ hay không.

### **2.5.Xây dựng hệ thống nhận diện mặt người đặc trưng**

Một hệ thống nhận diện khuôn mặt người có thể được xây dựng qua rất nhiều công đoạn khác nhau và rất phức tạp, nhưng ta có thể khái quát chung mọi hệ thống nhận dạng khuôn mặt người gồm có 3 bước cơ bản sau:





Hình 3. Hệ thống nhận diện mặt người đặc trưng

- *Tiền xử lý*: hệ thống nhận vào một ảnh tĩnh, sau đó có thể xử lý ảnh cho chất lượng tốt hơn, như chỉnh lại độ sáng, giảm độ nhiễu .. điều này giúp cho quá trình rút trích các đặc trưng của khuôn mặt được dễ dàng hơn rất nhiều. Quá trình tiền xử lý thường khá đơn giản và nhanh gọn nên không cần dùng các thuật toán phức tạp và mất nhiều thời gian.
- *Trích rút các đặc trưng*: 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.
- *Nhận diện khuôn mặt*: sau khi đã rút trích các đặc trưng, sẽ đưa vào khối nhận dạng để phân lớp đối tượng.

### 2.6. Một số phương pháp nhận diện mặt người

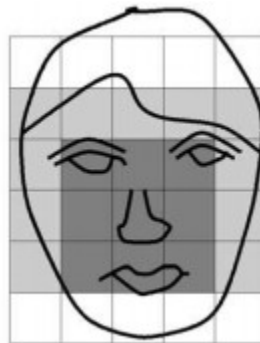
Hiện tại có khá nhiều phương pháp nhằm xác định khuôn mặt người trong ảnh, từ ảnh có chất lượng không tốt cho đến tốt, không màu (đen trắng) đến có màu. Nhưng có thể phân chia các phương pháp này thành bốn hướng tiếp cận chính :

- *Hướng tiếp cận dựa trên tri thức*: Mã hóa các hiểu biết của con người về khuôn mặt thành các luật. các luật này thường mô tả quan hệ của các đặc trưng của khuôn mặt.
- *Hướng tiếp cận dựa trên diện mạo*: Hướng tiếp cận này dùng một tập ảnh huấn luyện cho trước, rồi sau đó hệ thống sẽ xác định khuôn mặt người, do đó nó còn có tên là: hướng tiếp cận dựa trên phương pháp học.
- *Hướng tiếp cận dựa trên so sánh khớp mẫu*: Dùng các mẫu chuẩn của khuôn mặt người để mô tả cho khuôn mặt hay các đặc trưng của khuôn mặt.

- *Hướng tiếp cận dựa trên các đặc trưng không thay đổi:* Các thuật toán đi tìm các đặc trưng mô tả cấu trúc khuôn mặt người mà các đặc trưng này sẽ không thay đổi khi tư thế khuôn mặt, điều kiện ánh sáng hay vị trí đặt máy ảnh thay đổi.

### 2.6.1. Dựa trên tri thức

Trong hướng tiếp cận này, các luật sẽ phụ thuộc rất lớn vào tri thức của những tác giả nghiên cứu về bài toán xác định khuôn mặt người. Đây là hướng tiếp cận dạng top-down. Dễ dàng xây dựng các luật cơ bản để mô tả các đặc trưng của khuôn mặt và các quan hệ tương ứng. Ví dụ, một khuôn mặt thường có hai mắt đối xứng nhau qua trục thẳng đứng ở giữa khuôn mặt và có một mũi, một miệng. Các quan hệ của các đặc trưng có thể được mô tả như quan hệ về khoảng cách và vị trí. Thông thường sẽ trích đặc trưng của khuôn mặt trước tiên để có được các ứng viên, sau đó các ứng viên này sẽ được xác định thông qua các luật để biết ứng viên nào là khuôn mặt và ứng viên nào không phải khuôn mặt. Thường áp dụng quá trình xác định để giảm số lượng xác định sai.



Hình 4: Một phương pháp xác định khuôn mặt điển theo hướng tiếp cận top-down. Các luật được xây dựng dựa vào tri thức của người nghiên cứu về các đặc trưng của khuôn mặt (ví dụ như cường độ phân phối và sự khác nhau) của các vùng trên khuôn mặt

Một vấn đề khá phức tạp khi dùng hướng tiếp cận này là làm sao chuyển từ tri thức con người sang các luật một cách hiệu quả. Nếu các luật này quá chi tiết thì khi xác

định có thể xác định thiếu các khuôn mặt có trong ảnh, vì những khuôn mặt này không thể thỏa mãn tất cả các luật đưa ra. Nhưng các luật tổng quát quá thì có thể chúng ta sẽ xác định lầm một vùng nào đó không phải là khuôn mặt mà lại xác định là khuôn mặt. Và cũng khó khăn mở rộng yêu cầu từ bài toán để xác định các khuôn mặt có nhiều tư thế khác nhau.

Hai tác giả Yang và Huang dùng một phương thức theo hướng tiếp cận như sau để xác các khuôn mặt. Hệ thống của hai tác giả này bao gồm ba mức luật. Ở mức cao nhất, dùng một khung cửa sổ quét trên ảnh và thông qua một tập luật để tìm các ứng viên có thể là khuôn mặt. Ở mức kế tiếp, hai ông dùng một tập luật để mô tả tổng quát hình dáng khuôn mặt. Còn ở mức cuối cùng lại dùng một tập luật khác để xem xét ở mức chi tiết các đặc trưng khuôn mặt. Một hệ thống đa độ phân giải có thứ tự được dùng để xác định. Các luật ở mức cao nhất để tìm ứng viên như: “vùng trung tâm khuôn mặt có bốn phần với một mức độ đều cơ bản”, “phần xung quanh bên trên của một khuôn mặt có một mức độ đều cơ bản”, và “mức độ khác nhau giữa các giá trị xám trung bình của phần trung tâm và phần bao bên trên là đáng kể”. Độ phân giải thấp nhất của ảnh dùng để tìm ứng viên khuôn mặt mà còn tìm ở các mức phân giải tốt hơn. Ở mức hai, xem xét biểu đồ histogram của các ứng viên để loại bớt ứng viên nào không phải là khuôn mặt, đồng thời dò ra cạnh bao xung quanh ứng viên. Ở mức cuối cùng, những ứng viên nào còn lại sẽ được xem xét các đặc trưng của khuôn mặt về mắt và miệng. Hai ông đã dùng một chiến lược “từ thô đến mịn” hay “làm rõ dần” để giảm số lượng tính toán trong xử lý. Mặc dù tỷ lệ chính xác chưa cao, nhưng đây là tiền đề cho nhiều nghiên cứu sau này

Hai tác giả Kotropoulos và Pitas đưa một phương pháp tương tự dùng trên độ phân giải thấp. Hai ông dùng phương pháp chiếu để xác định các đặc trưng khuôn mặt. Kanade đã thành công với phương pháp chiếu để xác định biên của khuôn mặt. Với  $I(x,y)$  là giá trị xám của một điểm trong ảnh có kích thước  $m \times n$  ở tại vị trí  $(x,y)$ , các hàm để chiếu ảnh theo phương ngang và thẳng đứng được định nghĩa như sau:

$$\begin{cases} HI(x) = \sum_{y=1}^n I(x, y) \\ VI(y) = \sum_{x=1}^m I(x, y) \end{cases} \quad (2.1)$$

Dựa trên biểu đồ hình chiếu ngang, có hai cực tiểu địa phương khi hai ông xét quá trình thay đổi độ dốc của HI, đó chính là cạnh bên trái và phải của hai bên đầu. Tương tự với hình chiếu dọc VI, các cực tiểu địa phương cũng cho ta biết vị trí miệng, đỉnh mũi, và hai mắt. Các đặc trưng này đủ để xác định khuôn mặt. Hình 5.a cho một ví dụ về cách xác định như trên. Cách xác định này có tỷ lệ xác định chính xác là 86.5% cho trường hợp chỉ có một khuôn mặt thẳng trong ảnh và hình nền không phức tạp. Nếu hình nền phức tạp như hình 5.b thì rất khó tìm. Còn nếu ảnh có nhiều khuôn mặt (hình 5.c) thì sẽ không xác định được.



Hình 5: Phương pháp chiếu:

- (a) Ảnh chỉ có một khuôn mặt và hình nền đơn giản;
- (b) Ảnh chỉ có một khuôn mặt và hình nền phức tạp;
- (c) Ảnh có nhiều khuôn mặt

### 2.6.2. Hướng tiếp cận dựa trên đặc trưng không thay đổi.

Đây là hướng tiếp cận theo kiểu bottom-up. Các tác giả cố gắng tìm các đặc trưng không thay đổi của khuôn mặt người để xác định khuôn mặt người. Dựa trên nhận xét thực tế, con người dễ dàng nhận biết các khuôn mặt và các đối tượng trong các tư thế khác nhau và điều kiện ánh sáng khác nhau, thì phải tồn tại các thuộc tính

hay đặc trưng không thay đổi. Có nhiều nghiên cứu đầu tiên xác định các đặc trưng khuôn mặt rồi chỉ ra có khuôn mặt trong ảnh hay không. Các đặc trưng như: lông mày, mắt, mũi, miệng, và đường viền của tóc được trích bằng phương pháp xác định cạnh. Trên cơ sở các đặc trưng này, xây dựng một mô hình thống kê để mô tả quan hệ của các đặc trưng này và xác định sự tồn tại của khuôn mặt trong ảnh. Một vấn đề của các thuật toán theo hướng tiếp cận đặc trưng cần phải điều chỉnh cho phù hợp điều kiện ánh sáng, nhiễu, và bị che khuất. Đôi khi bóng của khuôn mặt sẽ tạo thêm cạnh mới, mà cạnh này lại rõ hơn cạnh thật sự của khuôn mặt, vì thế nếu dùng cạnh để xác định sẽ gặp khó khăn.

### a. Các đặc trưng khuôn mặt :

Sirohey đưa một phương pháp xác định khuôn mặt từ một ảnh có hình nền phức tạp. Phương pháp dựa trên cạnh (dùng phương pháp Candy và heuristics để loại bỏ các cạnh để còn lại duy nhất một đường bao xung quanh khuôn mặt. Một hình ellipse dùng để bao khuôn mặt, tách biệt vùng đầu và hình nền. Tỷ lệ chính xác của thuật toán là 80%. Cũng dùng phương pháp cạnh như Sirohey, Chetverikov và Lerch dùng một phương pháp dựa trên blob và streak (hình dạng giọt nước và sọc xen kẽ), để xác định theo hướng các cạnh. Hai ông dùng hai blob tối và ba blob sáng để mô tả hai mắt, hai bên gò má, và mũi. Mô hình này dùng các streak để mô tả hình dáng ngoài của khuôn mặt, lông mày, và môi. Dùng ảnh có độ phân giải thấp theo biến đổi Laplace để xác định khuôn mặt thông qua blob.

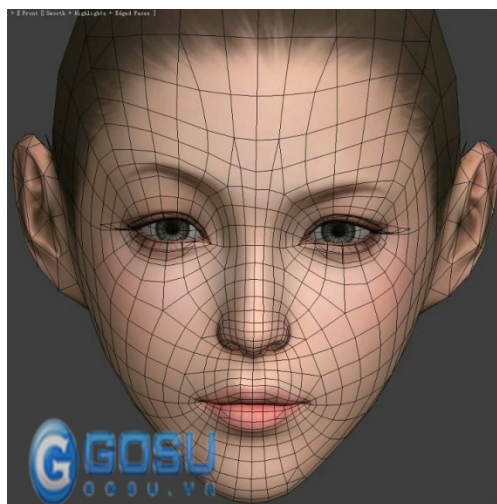
Leung trình bày một mô hình xác suất để xác định khuôn mặt ở trong ảnh có hình nền phức tạp trên cơ sở một bộ xác định đặc trưng cục bộ và so khớp đồ thị ngẫu nhiên. Dùng năm đặc trưng (hai mắt, hai lỗ mũi, phần nối giữa mũi và miệng) để mô tả một khuôn mặt. Luôn tính quan hệ khoảng cách với các đặc trưng cặp (như mắt trái, mắt phải), dùng phân bố Gauss để mô hình hóa. Một mẫu khuôn mặt được đưa ra thông qua trung bình tương ứng cho một tập đa hướng, đa tỷ lệ của bộ lọc đạo hàm Gauss. Từ một ảnh, các đặc trưng ứng viên được xác định bằng cách so khớp từng điểm ảnh khi lọc tương ứng với vector mẫu (tương tự mối tương quan), chọn hai ứng viên đặc trưng đứng đầu để tìm kiếm cho các đặc trưng khác của khuôn mặt. Giống

nghư xây dựng mộ đồ thị quan hệ mỗi node của đồ thị tương ứng như các đặc trưng của mộ khuôn mặt, đưa xác suất vào để xác định. Tỷ lệ xác định chính xác là 86%.

### b.Kết cấu khuôn mặt:

Khuôn mặt con người có những kết cấu riêng biệt mà có thể dùng để phân loại so với các đối tượng khác. Augusteijn và Skufca cho rằng hình dạng của khuôn mặt dùng làm kết cấu phân loại, gọi là kết cấu giống khuôn mặt (face-like texture). Tính kết cấu qua các đặc trưng thông kê thứ tự thứ hai (SGLD) trên vùng có kích thước  $16 \times 16$  điểm ảnh. Có ba loại đặc trưng được xem xét: màu da, tóc, và những thứ khác. Hai ông dùng mạng neural về mối tương quan cascade cho phân loại có giám sát các kết cấu và một ánh xạ đặc trưng tự tổ chức Kohonen để gom nhóm các lớp kết cấu khác nhau. Hai tác giả đề xuất dùng phương pháp bầu cử khi không quyết định được kết cấu đưa vào là kết cấu của da hay kết cấu của tóc.

Manian và Ross dùng biến đổi wavelet để xây dựng tập dữ liệu kết cấu của khuôn mặt trong ảnh xám thông qua nhiều độ phân giải khác nhau kết hợp xác suất thống kê để xác định khuôn mặt người. Mỗi mẫu sẽ có chín đặc trưng. Tỷ lệ chính xác là 87%, tỷ lệ xác định sai là 18%.



Hình 6. Kết cấu khuôn mặt

### c.Sắc màu của da:



Thông thường các ảnh màu không xác định trực tiếp trên toàn bộ dữ liệu ảnh mà các tác giả dùng tính chất sắc màu của da người (khuôn mặt người) để chọn ra được các ứng viên có thể là khuôn mặt người (lúc này dữ liệu đã thu hẹp đáng kể) để xác định khuôn mặt người. Tôi sẽ trình bày chi tiết về mô hình hóa màu da người ở một bài sau.



Hình 7. Màu sắc da mặt

### d. Đa đặc trưng:

Gần đây có nhiều nghiên cứu sử dụng các đặc trưng toàn cục như: màu da người, kích thước, và hình dáng để tìm các ứng viên khuôn mặt, rồi sau đó sẽ xác định ứng viên nào là khuôn mặt thông qua dùng các đặc trưng cục bộ (chi tiết) như: mắt, lông mày, mũi, miệng, và tóc. Tùy mỗi tác giả sẽ sử dụng tập đặc trưng khác nhau.

### **2.6.3. Hướng tiếp cận dựa trên so khớp mẫu**

Trong so khớp mẫu, các mẫu chuẩn của khuôn mặt (thường là khuôn mặt được chụp thẳng) sẽ được xác định trước hoặc xác định các tham số thông qua một hàm. Từ một ảnh đưa vào, tính các giá trị tương quan so với các mẫu chuẩn về đường viền khuôn mặt, mắt, mũi và miệng. Thông qua các giá trị tương quan này mà các tác giả quyết định có hay không có tồn tại khuôn mặt trong ảnh. Hướng tiếp cận này có lợi thế là rất dễ cài đặt, nhưng không hiệu quả khi tỷ lệ, tư thế, và hình dáng thay đổi. Nhiều



độ phân giải, đa tỷ lệ, các mẫu con, và các mẫu biến dạng được xem xét thành bất biến về tỷ lệ và hình dáng.

▪ ***Hướng tiếp cận này có thể đi theo phương pháp sau:***

a.Xác định mẫu trước:

Tác giả Sakai đã cố gắng thử xác định khuôn mặt người chụp thẳng trong ảnh. Ông dùng vài mẫu con về mắt, mũi, miệng, và đường viền khuôn mặt để mô hình hóa một khuôn mặt. Mỗi mẫu con được định nghĩa trong giới hạn của các đoạn thẳng. Các đường thẳng trong ảnh được trích bằng phương pháp xem xét thay đổi gradient nhiều nhất và so khớp các mẫu con. Đầu tiên tìm các ứng viên thông qua mối tương quan giữa các ảnh con và các mẫu về đường viền. Sau đó, so khớp với các mẫu con khác. Hay nói một cách khác, giai đoạn đầu xem như là giai đoạn sơ chế để tìm ứng viên, giai đoạn thứ hai là giai đoạn tinh chế để xác định có tồn tại hay không một khuôn mặt người. Ý tưởng này được duy trì cho đến các nghiên cứu sau này.

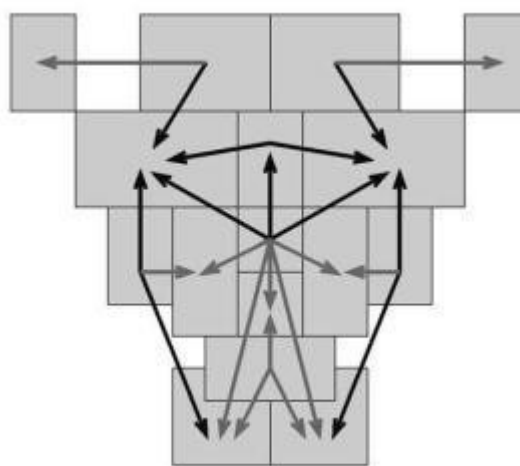
Còn tác giả Craw đưa ra một phương pháp xác định khuôn mặt người dựa vào các mẫu về hình dáng của các ảnh được chụp thẳng (dùng vẽ bề ngoài của hình dáng khuôn mặt). Đầu tiên dùng phép lọc Sobel để tìm các cạnh. Các cạnh này sẽ được nhóm lại theo một số ràng buộc. Sau đó, tìm đường viền của đầu, quá trình tương tự được lặp đi lặp lại với mỗi tỷ lệ khác nhau để xác định các đặc trưng khác như: mắt, lông mày, và môi.

Sau đó Craw mô tả một phương thức xác định dùng một tập có 40 mẫu để tìm các đặc trưng khuôn mặt và điều khiển chiến lược dò tìm.

Một số nhà nghiên cứu sử dụng hình chiếu như các mẫu để xác định khuôn mặt người. Silhouettes dùng PCA (phân tích thành phần chính – Principal Component Analysis - PCA) để có một tập hình chiếu cơ bản từ các mẫu khuôn mặt, hình chiếu được mô tả như một mảng các bit. Dùng đặc trưng hình chiếu riêng kết hợp biến đổi Hough để xác định khuôn mặt người. Sau đó một phương pháp xác định dựa trên đa loại mẫu để xác định các thành phần của khuôn mặt được trình bày. Phương pháp này định nghĩa một số giả thuyết để mô tả các khả năng của các đặc trưng khuôn mặt. Với

một khuôn mặt sẽ có một tập giả thuyết, lý thuyết Dempster-Shafer. Dùng một nhân tố tin cậy để kiểm tra sự tồn tại hay không của các đặc trưng của khuôn mặt, và kết hợp nhân tố tin cậy này với một độ đo để xem xét có hay không có khuôn mặt trong ảnh.

Sinha dùng một tập nhỏ các ảnh bất biến trong không gian ảnh để mô tả không gian các mẫu ảnh [15, 16]. Tư tưởng chính của ông dựa vào sự thay đổi mức độ sáng của các vùng khác nhau của khuôn mặt (như hai mắt, hai má, và trán), (quan hệ về mức độ sáng của các vùng còn lại thay đổi không đáng kể). Sau đây xác định các cặp tỷ số của mức độ sáng của một số vùng (một vùng tối hơn hay sáng hơn) cho ta một lượng bất biến khá hiệu quả. Ông lưu sự thay đổi độ sáng của các vùng trên khuôn mặt trong một tập thích hợp với các cặp quan hệ sáng hơn – tối hơn giữa các vùng nhỏ. Một khuôn mặt được xác định nếu nó thỏa mãn tất cả các cặp sáng hơn – tối hơn. Ý tưởng này xuất phát từ sự khác biệt của cường độ giữa các vùng cục bộ trên ảnh, sau này nó được mở rộng trên cơ sở biến đổi wavelet để xác định người đi bộ, xác định xe hơi, xác định khuôn mặt. Ý tưởng của Sinha còn được áp dụng trong các nghiên cứu về thị giác của robot. Hình 3 cho thấy mẫu khuôn mặt với 16 vùng và 23 quan hệ. Các quan hệ này được dùng để phân loại, có 11 quan hệ thiết yếu (các mũi tên màu đen) và 12 quan hệ xác thực (các mũi tên màu xám). Mỗi mũi tên là một quan hệ. Một quan hệ của hai vùng thỏa mãn mẫu khuôn mặt khi tỷ lệ giữa hai vùng vượt qua một ngưỡng và một khuôn mặt được xác định khi có 23 quan hệ thỏa mãn điều kiện trên.



Hình 8: Một mẫu khuôn mặt có kích thước 14 x 16 pixel được sử dụng trong phương pháp định

vị khuôn mặt của Sinha. Mẫu gồm 16 vùng trên khuôn mặt và 23 quan hệ (các mũi tên).

Một số phương pháp khác như của Froba và Zink lọc cạnh ở độ phân giải thấp rồi dùng biến đổi Hough để so khớp mẫu theo hướng cạnh để xác định hình dáng khuôn mặt ở dạng chụp hình thẳng ở dạng xám. Tỷ lệ chính xác trên 91%. Ngoài ra Shu và Jain còn xây dựng ngữ nghĩa khuôn mặt. Ngữ nghĩa xây dựng theo hình dáng và vị trí các thành phần khuôn mặt. Hai ông từ bộ ngữ nghĩa này tạo một đồ thị quan hệ để dễ dàng so khớp khi xác định khuôn mặt người ...

### b. Các mẫu bị biến dạng:

Tác giả Yuille dùng các mẫu biến dạng để mô hình hóa các đặc trưng của khuôn mặt, mô hình này có khả năng linh hoạt cho các đặc trưng khuôn mặt. Trong hướng tiếp cận này, các đặc trưng khuôn mặt được mô tả bằng các mẫu được tham số hóa. Một hàm năng lượng (giá trị) được định nghĩa để liên kết các cạnh, đỉnh, và thung lũng trong ảnh để tương ứng với các tham số trong mẫu. Mô hình này tốt nhất khi tối thiểu hàm năng lượng qua các tham số, Mặc dù kết quả tốt với mẫu biến dạng trong theo vết đối tượng trên đặc trưng không mô hình theo lưới, một hạn chế của hướng tiếp cận này là các mẫu biến dạng phải được khởi tạo trong phạm vi gần các đối tượng để xác định.

Hai tác giả Huang và Su dùng lý thuyết dòng chảy để xác định đường viền khuôn mặt dựa trên đặc tính hình học. Hai ông dùng lý thuyết tập đồng mức (Level Set) để loang từ các khởi động ban đầu để có được các khuôn mặt người.

### **2.6.4. Hướng tiếp cận dựa trên diện mạo**

#### a. Tư tưởng của hướng tiếp cận dựa trên diện mạo:

Trái ngược với các phương pháp so khớp mẫu với các mẫu đã được định nghĩa trước bởi những chuyên gia, các mẫu trong hướng tiếp cận này được học từ các ảnh mẫu. Một cách tổng quát, các phương pháp theo hướng tiếp cận này áp dụng các kỹ thuật theo hướng xác suất thống kê và máy học để tìm những đặc tính liên quan của khuôn mặt và không phải là khuôn mặt. Các đặc tính đã được học ở trong hình thái các mô hình phân bố hay các hàm biệt số nên dùng có thể dùng các đặc tính này để xác

định khuôn mặt người. Đồng thời, bài toán giảm số chiều thường được quan tâm để tăng hiệu quả tính toán cũng như hiệu quả xác định.

Có nhiều phương pháp áp dụng xác suất thống kê để giải quyết. Một ảnh hay một vector đặc trưng xuất phát từ một ảnh được xem như một biến ngẫu nhiên  $x$ , và biến ngẫu nhiên có đặc tính là khuôn mặt hay không phải khuôn mặt bởi công thức tính theo các hàm mật độ phân lớp theo điều kiện.

$P(x \mid \text{khuôn mặt})$  và  $P(x \mid \sim \text{khuôn mặt})$

Có thể dùng phân loại Bayes hoặc khả năng cực đại để phân loại một ứng viên là khuôn mặt hay không phải là khuôn mặt. Không thể cài đặt trực tiếp phân loại Bayes bởi vì số chiều của  $x$  khá cao, bởi vì  $P(x \mid \text{khuôn mặt})$  và  $P(x \mid \text{không phải khuôn mặt})$  là đa phương thức, và chưa thể hiểu nếu xây dựng các dạng tham số hóa một cách tự nhiên cho  $P(x \mid \text{khuôn mặt})$  và  $P(x \mid \text{không phải khuôn mặt})$ . Có khá nhiều nghiên cứu theo hướng tiếp cận này quan tâm xấp xỉ có tham số hay không có tham số cho  $P(x \mid \text{khuôn mặt})$  và  $P(x \mid \text{không phải khuôn mặt})$ .

Các tiếp cận khác trong hướng tiếp cận dựa trên diện mạo là tìm một hàm biệt số (như: mặt phẳng quyết định, siêu phẳng để tách dữ liệu, hàm ngưỡng) để phân biệt hai lớp dữ liệu: khuôn mặt và không phải khuôn mặt. Bình thường, các mẫu ảnh được chiếu vào không gian có số chiều thấp hơn, rồi sau đó dùng một hàm biệt số (dựa trên các độ đo khoảng cách) để phân loại, hoặc xây dựng mặt quyết định phi tuyến bằng mạng neural đa tầng. Hoặc dùng SVM (Support Vector Machine) và các phương thức kernel, chiếu hoàn toàn các mẫu vào không gian có số chiều cao hơn để dữ liệu bị rời rạc hoàn toàn và ta có thể dùng một mặt phẳng quyết định phân loại các mẫu khuôn mặt và không phải khuôn mặt.

### b. Một số phương pháp cụ thể trong hướng tiếp cận dựa trên diện mạo:

Có rất nhiều phương pháp học máy được sử dụng trong hướng tiếp cận dựa trên diện mạo. Trong phần này chỉ nêu ra một số phương pháp cùng những nghiên cứu liên quan đến phương pháp đấy. Riêng phương pháp adaboost sẽ được trình bày trong phần sau của báo cáo.

### ❖ b1. Adaboost:

Adaboost được đánh giá là phương pháp tiếp cận nhanh nhất trong các thuật toán học máy. Nó thường được kết hợp với các mô hình cascade of classifiers để tăng tốc độ phát hiện khuôn mặt trong ảnh. Tư tưởng của thuật toán adaboost đây là kết hợp các bộ phân loại yếu thành một bộ phân loại mạnh. Trong quá trình xây dựng, bộ phân loại yếu tiếp theo sẽ được xây dựng dựa trên các đánh giá về các bộ phân loại yếu trước, cuối cùng các bộ phân loại yếu sẽ được kết hợp để trở thành bộ phân loại mạnh.

Viola và Jones dùng AdaBoost kết hợp cascade để xác định khuôn mặt người [17] với các đặc trưng dạng Haar-like. Tốc độ xử lý khá nhanh và tỷ lệ chính xác hơn 80%. Ngoài ra Schneiderman và Kanade dùng wavelet để trích đặc trưng, sau đây cũng xây dựng hệ thống học với Adaboost, dựa trên xác suất để xác định khuôn mặt người. Tỷ lệ chính xác trên của phương pháp này lên đến 90%.

### ❖ b2. Mạng Neural:

Về cơ bản mạng Neural là một mạng các phần tử (gọi là neuron) kết nối với nhau thông qua các liên kết (các liên kết này được gọi là trọng số liên kết) để thực hiện một công việc cụ thể nào đó. Khả năng xử lý của mạng neuron được hình thành thông qua quá trình hiệu chỉnh trọng số liên kết giữa các neuron, nói cách khác là học từ tập hợp các mẫu huấn luyện.

Mạng Neural rất hay được sử dụng trong các bài toán nhận dạng mẫu, nhận dạng mặt người cũng có thể xem là một bài toán nhận dạng với hai loại mẫu (mặt người hoặc không phải mặt người) và như vậy có thể sử dụng mạng Neural khá hiệu quả kể cả khi các mẫu khuôn mặt có độ phức tạp cao. Tuy nhiên, một điều trở ngại đó là các kiến trúc mạng đều có tính tổng quát cao, do đó, khi áp dụng ta phải tính toán rõ số lượng tầng, số lượng node, tỷ lệ học ... cho từng trường hợp cụ thể.

### ❖ b3. Support Vector Machine:

Support Vector Machine ( SVM ) được Vladimir Vapnik đưa ra vào năm 1995 để giải quyết vấn đề nhận dạng mẫu hai lớp sử dụng nguyên tắc cực tiểu hóa rủi ro cấu trúc (Structural Risk Minimization). Đây là phương pháp tiếp cận phân loại văn bản rất hiệu quả. Ưu điểm của SVM là giải thuật này được xây dựng trên ý tưởng cực tiểu rủi

ro cấu trúc. Nguồn gốc của SVM dựa trên sự chắc chắn về lỗi chính xác, có thể phân loại ngẫu nhiên các mẫu đối tượng được chọn mà lỗi được giữ sao cho nhỏ nhất. Vì vậy, giải thuật SVM giúp giảm thiểu biên trên các lỗi chính xác và làm cho hệ thống tin cậy hơn

Cho trước một tập huấn luyện, các ảnh được biểu diễn dưới dạng vector. Trong không gian vector, mỗi vector được biểu diễn bởi một điểm. Phương pháp SVM sẽ tìm một siêu phẳng quyết định để phân chia không gian vector thành hai lớp (khuôn mặt và không phải khuôn mặt). Chất lượng của siêu phẳng này phụ thuộc vào khoảng cách giữa các vector, tức là phụ thuộc vào các đặc trưng của ảnh.

Support Vector Machine đã được Osuna [5] áp dụng đầu tiên (để xác định khuôn mặt người). SVM được xem như là một kiểu phân loại mới vì trong khi hầu hết các phương pháp phân loại khác (như Mạng Bayes, Neural, RBF) đều dùng tiêu chí tối thiểu lỗi huấn luyện (rủi ro do kinh nghiệm), trong khi SVM dùng quy nạp (được gọi là tối thiểu rủi ro cấu trúc). Phân loại SVM là một phân loại tuyến tính, do đó nó cũng dùng một siêu phẳng để tách dữ liệu. Dựa trên một số kết hợp có các trọng số của một tập con nhỏ các vector huấn luyện, các vector này được gọi là support vector. Ước lượng siêu phẳng trong SVM thì tương đương giải một bài toán tuyến tính bậc hai. Osuna [5] đã phát triển phương pháp SVM một cách hiệu quả cho bài toán xác định khuôn mặt người. Ông dùng 10,000,000 mẫu có kích thước 19x19 pixel, hệ thống của ông có tỷ lệ lỗi ít hơn và nhanh hơn rất nhiều.

#### ❖ b4. Mô hình Markov ẩn:

Mô hình Markov ẩn ( Hidden Markov Model – HMM ) là một mô hình thống kê trong đó hệ thống được mô hình hóa được cho là một quá trình Markov với các tham số không biết trước và nhiệm vụ là xác định các tham số ẩn từ các tham số quan sát được, dựa trên sự thừa nhận này. Các tham số của mô hình được rút ra sau đó có thể sử dụng để thực hiện các phân tích kế tiếp, ví dụ cho các ứng dụng nhận dạng mẫu.

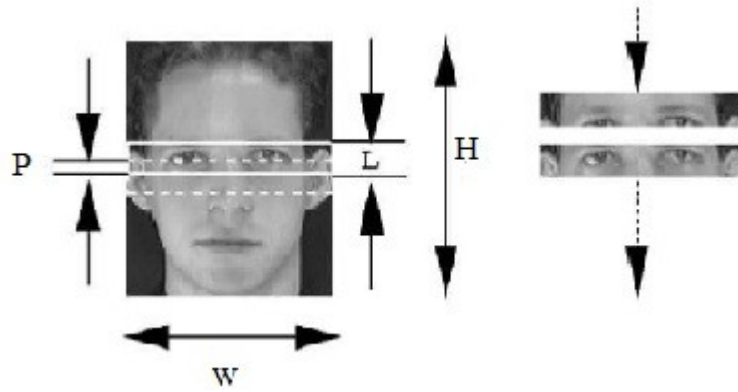
Một giả thuyết quan trọng của mô hình Markov ẩn là các mẫu có thể được đặc tính hóa như các tiến trình ngẫu nhiên có tham số và các tham số này được ước lượng chính xác, đây là một trong những định nghĩa rõ ràng. Khi phát triển HMM để giải quyết bài toán nhận dạng mẫu, phải xác định rõ có bao nhiêu trạng thái ẩn đầu tiên cho

hình thái mô hình. Sau đó, huấn luyện HMM học xác suất chuyển tiếp giữa các trạng thái từ các mẫu, mà mỗi mẫu được mô tả như một chuỗi các quan sát. Mục tiêu huấn luyện HMM là cực đại hóa xác suất của quan sát từ dữ liệu huấn luyện bằng cách điều chỉnh các tham số trong mô hình HMM thông qua phương pháp phân đoạn Viterbi chuẩn và các thuật toán Baum-Welch. Sau khi huấn luyện xong, dựa vào xác suất để xác định một quan sát thuộc lớp nào.

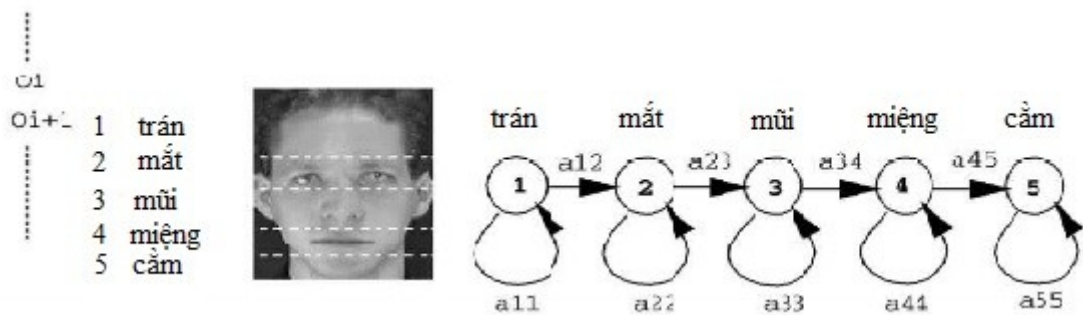
Để hình dung về HMM, ta xét ví dụ cụ thể sau: Giả sử A có một người bạn B sống ở rất xa. Hàng ngày thì B đều gọi điện cho A để kể về những việc mà B đã làm trong ngày. B chỉ có 3 việc mà anh ta thích làm đó là: việc thứ nhất là đi dạo, thứ hai là đi chợ và thứ ba là dọn phòng. Và việc anh ta làm việc gì trong ngày phụ thuộc rất lớn vào thời tiết của ngày hôm đấy. Như vậy, dù không nhận được thông tin cụ thể về thời tiết trong ngày ở nơi người B ở, nhưng người vẫn có thể dựa vào việc B làm trong ngày để đoán về thời tiết hôm đấy. Như vậy nếu coi thời tiết chỉ có hai trạng thái là nắng và mưa thì thời tiết là một chuỗi Markov cụ thể, và nó là ẩn đối với người A. Dữ liệu quan sát được ở đây là việc làm trong ngày của người B. Toàn bộ hệ thống này là một mô hình Markov ẩn. Ví dụ này được xem xét kỹ hơn trong thuật toán Viterbi.

Một giả thuyết quan trọng của mô hình Markov ẩn là các mẫu có thể được xem như các tiến trình ngẫu nhiên có tham số và các tham số này được ước lượng chính xác. Khi phát triển HMM để giải quyết bài toán nhận dạng mẫu, phải xác định rõ có bao nhiêu trạng thái ẩn đầu tiên cho hình thái mô hình. Sau đó, huấn luyện HMM học xác suất chuyển tiếp giữa các trạng thái từ các mẫu, mà mỗi mẫu được mô tả như một chuỗi các quan sát. Mục tiêu huấn luyện HMM là cực đại hóa xác suất của quan sát từ dữ liệu huấn luyện bằng cách điều chỉnh các tham số trong mô hình HMM thông qua phương pháp phân đoạn Viterbi chuẩn và các thuật toán Baum-Welch. Sau khi huấn luyện xong, dựa vào xác suất để xác định một quan sát thuộc lớp nào.



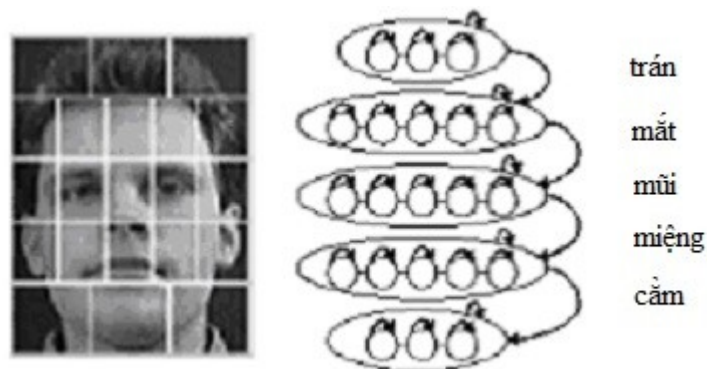


Hình 9. Các vector quan sát để huấn luyện cho HMM



Hình 10. Các trạng thái ẩn

Đối với bài toán phát hiện và nhận dạng mặt người, ta chia khuôn mặt thành các vùng khác nhau như đầu, mắt, mũi, miệng, và cằm. Có thể nhận dạng một mẫu khuôn mặt người bằng cách thực hiện tiến trình xem xét các vùng quan sát theo một thứ tự thích hợp (từ trên xuống dưới, từ trái qua phải). Mục tiêu của hướng tiếp cận này là kết hợp các vùng đặc trưng khuôn mặt với các trạng thái của mô hình.



Hình 11. Xác định khuôn mặt bằng HMM, mỗi trạng thái lại có những trạng thái nhỏ bên trong: trạng thái trán có ba trạng thái nhỏ bên trong; trạng thái mắt có năm trạng thái nhỏ bên trong.



Thường các phương pháp dựa vào HMM sẽ xem xét một mẫu khuôn mặt như một chuỗi các vector quan sát, với mỗi vector là một dãy các điểm ảnh (hình 4 và hình 6). Trong quá trình huấn luyện và kiểm tra, một ảnh được quét theo một thứ tự và một quan sát được xem như một khối các điểm ảnh (hình 4 và hình 6). Sau đấy, áp dụng một định hướng theo xác suất để chuyển từ trạng thái này sang trạng thái khác ( hình 5 ), dữ liệu ảnh được mô hình hóa bằng phân bố Gauss đa biến. Một chuỗi quan sát bao gồm tất cả giá trị cường độ từ mỗi khối. Kết quả xuất ra cho biết quan sát thuộc lớp nào.

Một nghiên cứu điển hình trong các tiếp cận này đó là của Samaria. Samaria đã dùng năm trạng thái tương ứng năm vùng như trong hình 6 để mô hình hóa tiến trình xác định khuôn mặt người. Ông huấn luyện từng vùng cho HMM. Mỗi trạng thái sẽ phụ trách xem xét một vùng tương ứng để đưa ra quyết định phù hợp. Nếu kết quả xem xét cuối cùng vượt qua một ngưỡng thì quan sát này sẽ là khuôn mặt người.

### ❖ b5. Phân loại Bayes:

Bayes là một bộ phân loại tuyến tính dựa trên xác suất. Tư tưởng của nó là dựa vào xác suất của các đặc trưng xuất hiện trong khuôn mặt. Đây là phương pháp đơn giản, cài đặt không phức tạp, tốc độ nhanh, với tập huấn luyện lớn thì cho kết quả vẫn tương đối chính xác.

Một nghiên cứu điển hình dùng phân loại Bayes vào xác định mặt người trong hướng tiếp cận dựa trên diện mạo đó là nghiên cứu của Schneiderman và Kanade. Hai ông đã mô tả một phân loại Naïve Bayes để ước lượng xác suất chung của diện mạo cục bộ và vị trí của các mẫu (các vùng trên khuôn mặt) ở nhiều độ phân giải khác nhau. Ứng với mỗi độ phân giải, khuôn mặt người được chia thành bốn vùng chữ nhật, các vùng này được chiếu xuống không gian có số chiều bé hơn và được lượng tử hóa thành tập các mẫu có giới hạn. Sau đấy thống kê các vùng đã được chiếu để mã hóa diện mạo cục bộ. Nếu xác suất lớn hơn xác suất đạt được lớn hơn xác suất tiền nghiệm thì kết luận có khuôn mặt người. Hướng tiếp cận này còn cho phép xác định các khuôn mặt bị xoay hoặc góc nhìn nghiêng.

### CHƯƠNG 3. CƠ SỞ LÝ THUYẾT

#### 3.1. Phát hiện mặt người

Học theo adaboost là một cách trong hướng tiếp cận dựa trên diện mạo, Viola và Jones dùng AdaBoost kết hợp cascade để xác định khuôn mặt người với các đặc trưng dạng Haar wavelet-like. Tốc độ xử lý khá nhanh và tỷ lệ chính xác hơn 80% trên ảnh xám.

Thuật toán học máy Adaboost được phát triển thuật toán boosting, do đó nhóm sẽ trình bày một chút về thuật toán boosting trước khi trình bày về adaboost.

### 3.1.1. Tiếp cận Boosting

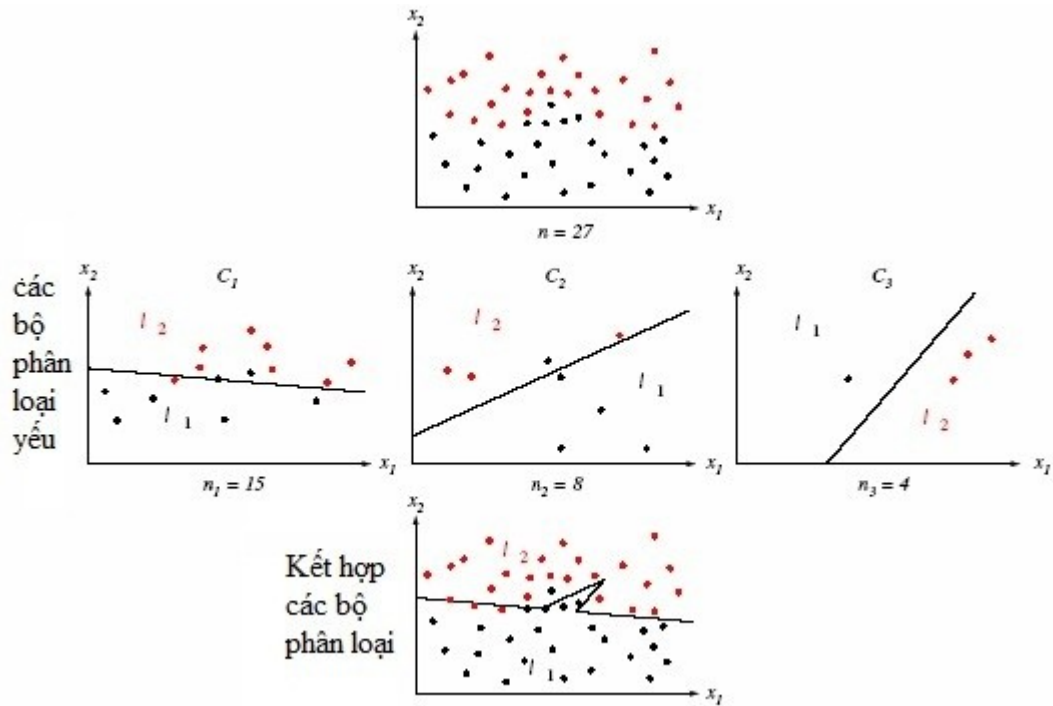
Về lịch sử, boosting bắt nguồn từ câu hỏi nổi tiếng được đưa ra bởi Kearns vào năm 1989 : “Liệu có thể tạo ra một strong classifier từ một tập các bộ phân loại yếu?”. Năm 1990, Robert Schapire đưa ra thuật toán boosting đầu tiên, tiếp đến năm 1993 thì nó được Drucker, Schapire và Simard kiểm nghiệm trong trong các chương trình nhận dạng ( OCR application ). Freund đã tiếp tục các nghiên cứu của Schapire, và đến năm 1995 thì ông cùng với Schapire phát triển boosting thành adaboost.

Như vậy, nguyên lý cơ bản của boosting là sự kết hợp các weak classifiers thành một strong classifier. Trong đó, weak classifier là các bộ phân loại đơn giản chỉ cần có độ chính xác trên 50%. Bằng cách này, chúng ta nói bộ phân loại đã được “boost”.

Để hiểu cách hoạt động của thuật toán boosting, ta xét một bài toán phân loại 2 lớp (mẫu cần nhận dạng chỉ thuộc một trong hai lớp) với  $D$  là tập huấn luyện gồm có  $n$  mẫu. Trước tiên, chúng ta sẽ chọn ngẫu nhiên ra  $n_1$  mẫu từ tập  $D$  ( $n_1 < n$ ) để tạo tập  $D_1$ . Sau đó, chúng ta sẽ xây dựng weak classifier đầu tiên  $C_1$  từ tập  $D_1$ . Tiếp theo, chúng ta xây dựng tập  $D_2$  để huấn luyện bộ phân loại  $C_2$ .  $D_2$  sẽ được xây dựng sao cho một nửa số mẫu của nó được phân loại đúng bởi  $C_1$  và nửa còn lại bị phân loại sai bởi  $C_1$ . Bằng cách này,  $D_2$  chứa đựng những thông tin bổ sung cho  $C_1$ . Bây giờ chúng ta sẽ xây huấn luyện  $C_2$  từ  $D_2$ .

Tiếp theo, chúng ta sẽ xây dựng tập  $D_3$  từ những mẫu không được phân loại tốt bởi sự kết hợp giữa  $C_1$  và  $C_2$ : những mẫu còn lại trong  $D$  mà  $C_1$  và  $C_2$  cho kết quả khác nhau. Như vậy,  $D_3$  sẽ gồm những mẫu mà  $C_1$  và  $C_2$  hoạt động không hiệu quả. Sau cùng, chúng ta sẽ huấn luyện bộ phân loại  $C_3$  từ  $D_3$ .

Bây giờ chúng ta đã có một strong classifier: sự kết hợp  $C_1$ ,  $C_2$  và  $C_3$ . Khi tiến hành nhận dạng một mẫu  $X$ , kết quả sẽ được quyết định bởi sự thỏa thuận của 3 bộ  $C_1$ ,  $C_2$  và  $C_3$ : Nếu cả  $C_1$  và  $C_2$  đều phân  $X$  vào cùng một lớp thì lớp này chính là kết quả phân loại của  $X$ ; ngược lại, nếu  $C_1$  và  $C_2$  phân  $X$  vào 2 lớp khác nhau,  $C_3$  sẽ quyết định  $X$  thuộc về lớp nào



Hình 12. Boosting

### 3.1.2. Adaboost

Như đã biết, AdaBoost ( Adaptive Boost ) là một bộ phân loại mạnh phi tuyến phức dựa trên hướng tiếp cận boosting được Freund và Schapire đưa ra vào năm 1995. Adaboost hoạt động trên nguyên tắc kết hợp tuyến tính các weak classifier dựa trên các đặc trưng Haar- Line để hình thành một strong classifier.

Để có thể kết hợp các bộ phân loại yếu, adaboost sử dụng một trọng số (weight) để đánh dấu các mẫu khó nhận dạng. Trong quá trình huấn luyện, cứ mỗi weak classifier được xây dựng, thuật toán sẽ tiến hành cập nhật lại trọng số để chuẩn bị cho việc xây dựng weak classifier tiếp theo: tăng trọng số của các mẫu bị nhận dạng sai và giảm trọng số của các mẫu được nhận dạng đúng bởi weak classifier vừa xây dựng. Bằng cách này, các weak classifier sau có thể tập trung vào các mẫu mà các weak classifier trước đó chưa thực hiện tốt. Sau cùng các weak classifier sẽ được kết hợp tùy theo mức độ ‘tốt’ của chúng để tạo nên một strong classifier.

Các weak classifiers  $h_k(x)$  là các bộ phân loại yếu, được biểu diễn như sau:

$$h_k(x) = \begin{cases} 1 & \text{nếu } p_k f_k(x) < p_k \theta_k \\ 0 & \text{nếu ngược lại} \end{cases} \quad \text{Trong đó:}$$

$X$  : mẫu hay cửa sổ con cần xét (  $X = (x_1, x_2, \dots, x_n)$  là vector đặc trưng của mẫu)

$O_k$ : ngưỡng ( $O = \theta_k$ )

$f_k$ : giá trị của đặc trưng Haar-like

$p_k$ : hệ số quyết định chiều của bất phương trình

Công thức trên có thể diễn giải như sau: Nếu giá trị đặc trưng của mẫu cho bởi hàm đánh giá của bộ phân loại vượt qua một ngưỡng cho trước thì mẫu đấy là khuôn mặt ( gọi là object : đối tượng cần nhận dạng ), ngược lại thì mẫu là background ( không phải là đối tượng ).

Thuật toán AdaBoost:

1. Cho một tập gồm  $n$  mẫu có đánh dấu  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  với  $x_k \in (x_{k1}, x_{k2}, \dots, x_{km})$  là vector đặc trưng và  $y_k \in (-1, 1)$  là nhãn của mẫu (1 ứng với *object*, -1 ứng với *background*).
2. Khởi tạo trọng số ban đầu cho tất cả các mẫu: với  $m$  là số mẫu đúng (ứng với *object* và  $y = 1$ ) và  $l$  là số mẫu sai (ứng với *background* và  $y = -1$ ).

$$w_{1,k} = \frac{1}{2m}, \frac{1}{2l}$$

3. Xây dựng  $T$  weak classifiers

Lặp  $t = 1, \dots, T$

- Với mỗi đặc trưng trong vector đặc trưng, xây dựng một weak classifier  $h_j$  với ngưỡng  $\theta_j$  và lỗi  $\epsilon_j$ .

$$w_{t,k} \vee h_j(x_k) - y_k \vee \epsilon_j$$

$$\epsilon_j = \sum_k^n \epsilon_j$$

- Chọn ra  $h_j$  với  $\epsilon_j$  nhỏ nhất, ta được  $h_t$ :

$$h_t: X \rightarrow \{1, -1\}$$

- Cập nhật lại trọng số:

$$w_{t+1,k} = \frac{w_{t,k}}{Z_t} \times \begin{cases} e^{-\alpha_t}, & h_t(x_k) = y_k \\ e^{\alpha_t}, & h_t(x_k) \neq y_k \end{cases}$$

- Trong đó:

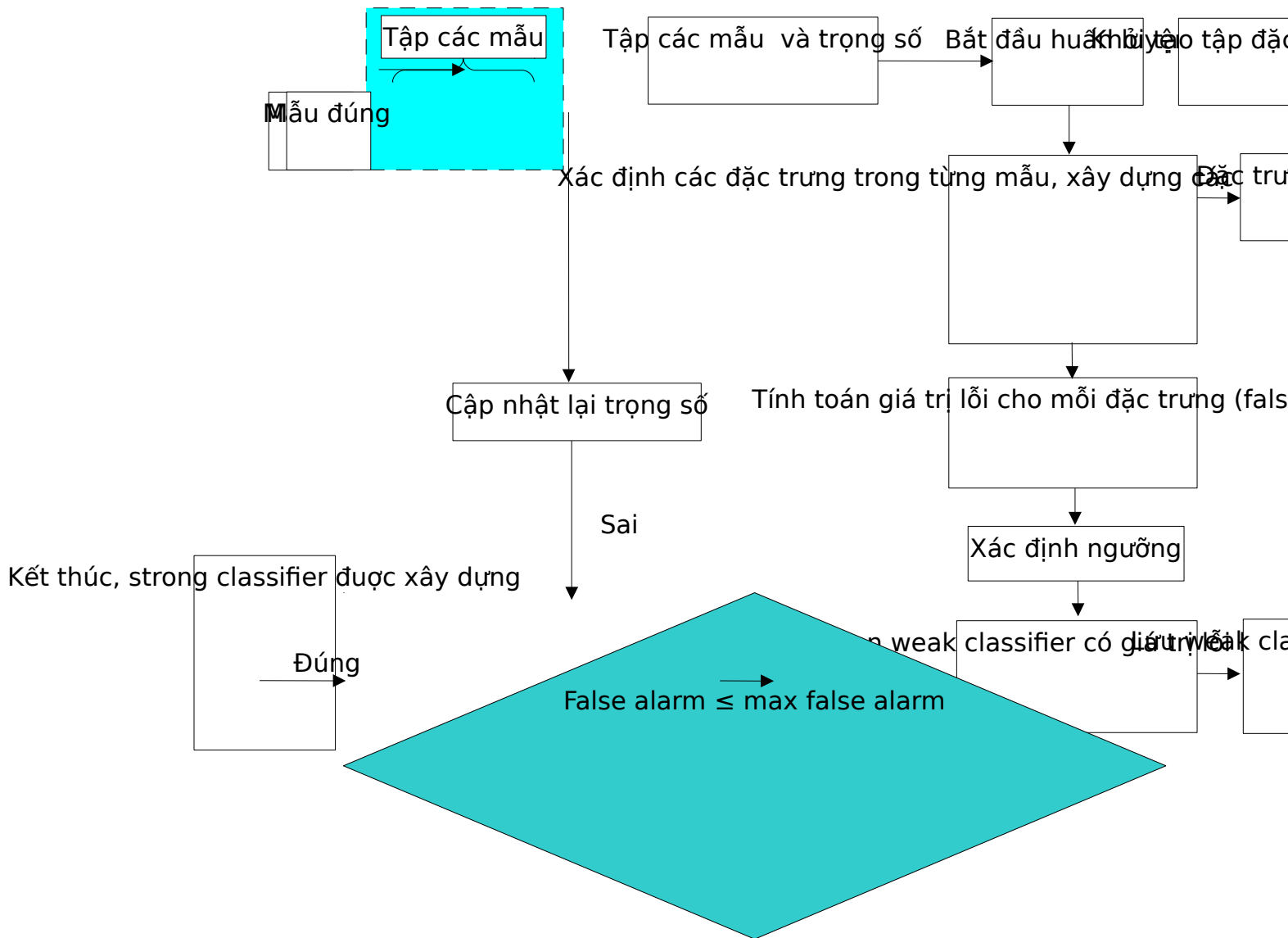
$$\alpha_t = \frac{1}{2} \ln \left( \frac{1 - \epsilon_j}{\epsilon_j} \right)$$

$Z_t$  : Hệ số dùng để đưa  $w_{t+1}$  về đoạn  $[0,1]$  (normalization factor)

4. Strong classifier được xây dựng :

$$H(x) = \text{đấu} \left( \sum_t \alpha_t h_t(x) \right)$$

Sơ đồ khối:



Hình 13. Sơ đồ khối thuật toán AdaBoost

Quá trình huấn luyện bộ phân loại được thực hiện bằng một vòng lặp mà ở mỗi bước lặp, thuật toán sẽ chọn ra weak classifier  $h_t$  thực hiện việc phân loại với lỗi  $\epsilon_t$  nhỏ nhất (do đó sẽ là bộ phân loại tốt nhất) để bổ sung vào strong classifier. Mỗi khi chọn được 1 bộ phân loại  $h_t$ , Adaboost sẽ tính giá trị  $\alpha_t$  theo công thức ở trên.  $\alpha_t$  cũng được chọn trên nguyên tắc làm giảm thiểu giá trị lỗi  $\epsilon_t$ .

Hệ số  $\alpha_t$  nói lên mức độ quan trọng của  $h_t$ :

- Trong công thức của bộ phân loại  $H(x)$ :

$$H(x) = \text{dấu} \left( \sum_{t=1}^T \alpha_t h_t(x) \right)$$

Ta thấy tất cả các bộ phân loại  $h_t$  đều có đóng góp vào kết quả của bộ phân loại  $H(x)$ , và mức độ đóng góp của chúng phụ thuộc vào giá trị  $\alpha_t$  tương ứng:  $h_t$  với  $\alpha_t$  càng lớn thì nó càng có vai trò quan trọng trong  $H(x)$ .

- Trong công thức tính  $\alpha_t$ :

$$\alpha_t = \frac{1}{2} \ln \left( \frac{1 - \varepsilon_j}{\varepsilon_j} \right)$$

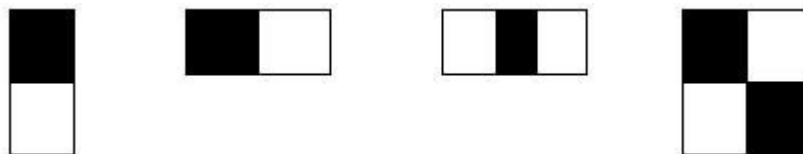
Dễ thấy giá trị  $\alpha_t$  tỉ lệ nghịch với  $\varepsilon_t$ . Bởi vì  $h_t$  được chọn với tiêu chí đạt  $\varepsilon_t$  nhỏ nhất, do đó nó sẽ đảm bảo giá trị  $\alpha_t$  lớn nhất.

Sau khi tính được giá trị  $\alpha_t$ , Adaboost tiến hành cập nhật lại trọng số của các mẫu: tăng trọng số các mẫu mà  $h_t$  phân loại sai, giảm trọng số các mẫu mà  $h_t$  phân loại đúng. Bằng cách này, trọng số của mẫu phản ánh được mức độ khó nhận dạng của mẫu đó và  $h_{t+1}$  sẽ ưu tiên học cách phân loại những mẫu này.

Vòng lặp xây dựng strong classifier sẽ dừng lại sau  $T$  lần lặp. Trong thực tế cài đặt (thư viện OpenCV của Intel), người ta ít sử dụng giá trị  $T$  vì không có công thức nào đảm bảo tính được giá trị  $T$  tối ưu cho quá trình huấn luyện. Thay vào đó, người ta sử dụng giá trị max false positive hay max false alarm (tỉ lệ nhận dạng sai tối đa các mẫu background). Tỉ lệ này của bộ phân loại cần xây dựng không được phép vượt quá giá trị này. Khi đó, qua các lần lặp, false alarm của strong classifier  $H_t(x)$  xây dựng được (tại lần lặp thứ  $t$ ) sẽ giảm dần, và vòng lặp kết thúc khi tỉ lệ này thấp hơn max false alarm.

### 3.1.3. Các đặc trưng Haar-Like

Viola và Jones dùng 4 đặc trưng cơ bản [20] để xác định khuôn mặt người. Mỗi đặc trưng Haar-like là sự kết hợp của hai hay ba hình chữ nhật "trắng" hay "đen" như trong hình sau:

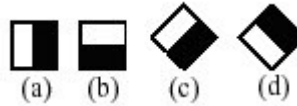


Hình 14. Đặc trưng Haar-like cơ bản

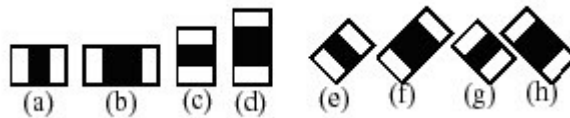
Để áp dụng các đặc trưng này vào việc bài toán xác định mặt người, 4 đặc trưng Haar-like cơ bản được mở rộng ra, và được chia làm 3 tập đặc trưng như sau:



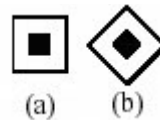
## 1. Đặc trưng cạnh (edge features):



## 2. Đặc trưng đường (line features):



## 3. Đặc trưng xung quanh tâm (center-surround features):



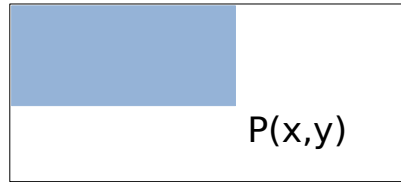
Hình 15. Các đặc trưng mở rộng của các đặc trưng Haar-like cơ sở

Lợi ích của các đặc trưng Haar-like là nó diễn đạt được tri thức về các đối tượng trong ảnh (bởi vì nó biểu diễn mối liên hệ giữa các bộ phận của đối tượng), điều mà bản thân từng điểm ảnh không diễn đạt được. Để tính giá trị các đặc trưng haar-like, ta tính sự chênh lệch giữa tổng của các pixel của các vùng đen và các vùng trắng như trong công thức sau:

$$f(x) = \text{Tổng}_{\text{vùng đen}}(\text{pixel}) - \text{Tổng}_{\text{vùng trắng}}(\text{pixel})$$

Như vậy ta có thể thấy rằng, để tính các giá trị của đặc trưng Haar-like, ta phải tính tổng của các vùng pixel trên ảnh. Nhưng để tính toán các giá trị của các đặc trưng Haar-like cho tất cả các vị trí trên ảnh đòi hỏi chi phí tính toán khá lớn, không đáp ứng được cho các ứng dụng đòi hỏi tính run-time. Do đó Viola và Jones đưa ra một khái niệm gọi là Integral Image để tính toán nhanh cho các đặc trưng cơ bản. Sau này, Lienhart kế thừa gọi Integral Image là SAT – Summed Area Table và đưa ra thêm khái niệm RSAT – Rotated Summed Area Table dùng để tính toán nhanh cho các đặc trưng xoay 1 góc 45°. Integral Image là một mảng 2 chiều với kích thước bằng với kích thước của ảnh cần tính các đặc trưng Haar-like, với mỗi phần tử của mảng này được tính bằng

cách tính tổng của điểm ảnh phía trên (dòng-1) và bên trái (cột-1) của nó. Bắt đầu từ vị trí trên, bên trái đến vị trí dưới, phải của ảnh, việc tính toán này đơn thuần chỉ dựa trên phép cộng số nguyên đơn giản, do đó tốc độ thực hiện rất nhanh.



$$P(x,y) = \sum_{x' \leq x, y' \leq y} i(x', y')$$

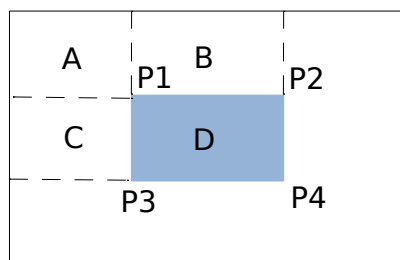
Hình 16. Cách tính Integral Image của ảnh

Sau khi đã tính được Integral Image, việc tính tổng điểm ảnh của một vùng bất kỳ nào đó trên ảnh thực hiện rất đơn giản theo cách sau: (Giả sử ta cần tính tổng điểm ảnh của vùng D như trong hình 11):

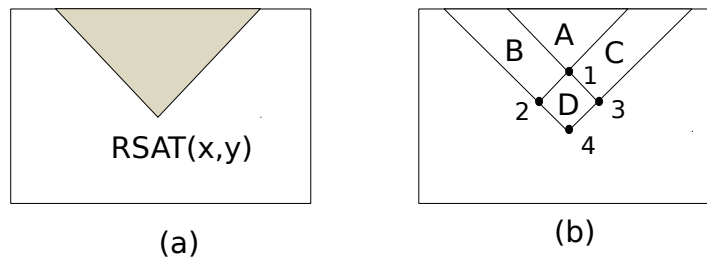
$$D = A + B + C + D - (A+B) - (A+C) + A$$

Với  $A + B + C + D$  chính là giá trị tại điểm P4 trên Integral Image, tương tự như vậy  $A+B$  là giá trị tại điểm P2,  $A+C$  là giá trị tại điểm P3, và  $A$  là giá trị tại điểm P1. Vậy ta có thể viết lại biểu thức tính D ở trên như sau:

$$D = (x_4, y_4) - (x_2, y_2) - (x_3, y_3) + (x_1, y_1)$$



Hình 17. Ví dụ cách tính nhanh tổng các điểm ảnh của vùng D trên ảnh



Hình 12: Ví dụ cách tính nhanh tổng điểm ảnh của vùng D trên ảnh với các đặc trưng xoay 45°

Với các đặc trưng Haar-like xoay 45° Integral Image tại một điểm (x, y) được tính theo công thức:

$$P(x, y) = \sum_{x' \leq x, y' \leq y} I(x', y')$$

Tổng pixel của một vùng bất kỳ trên ảnh vẫn được tính theo cách sau:

$$D = A + B + C + D - (A+B) - (A+C) + A$$

Như vậy tổng các điểm ảnh trong một hình chữ nhật (kể cả trường hợp xoay 45°) bất kì đều có thể được tính nhanh dựa trên integral image tại 4 đỉnh của nó :

$$\text{Sum}(D) = .4 - 2 - 3 + 1$$

### 3.1.4.Cascade of Classifiers

Ta thấy quá trình huấn luyện, bộ phân loại phải duyệt qua tất cả các đặc trưng của các mẫu trong tập training. Việc này tốn rất nhiều thời gian. Tuy nhiên, trong các mẫu đưa vào, không phải mẫu nào cũng thuộc loại khó nhận dạng, có những mẫu background rất dễ nhận ra (ta gọi đây là những mẫu background đơn giản). Đối với những mẫu này, ta chỉ cần xét một hay vài đặc trưng đơn giản là có thể nhận diện được chứ không cần xét tất cả các đặc trưng. Nhưng đối với các bộ phân loại thông thường thì cho dù mẫu cần nhận dạng là dễ hay khó thì nó vẫn sẽ xét tất cả các đặc trưng mà nó rút ra được trong quá trình học. Do đó, chúng tốn thời gian xử lý một cách không cần thiết.

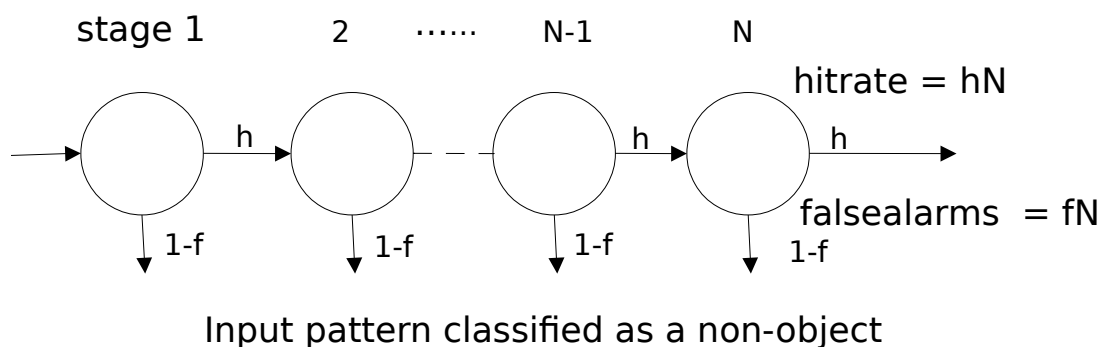
Cascade of Classifiers được xây dựng chính là nhằm rút ngắn thời gian xử lý, giảm thiểu false alarm cho bộ phân loại. Cascade tree gồm nhiều stage (hay còn gọi là layer), mỗi stage của cây sẽ là một stage classifier. Một mẫu để được phân loại là đối tượng thì nó cần phải đi qua hết tất cả các stages của cây. Các stage classifiers ở stage sau được huấn luyện bằng những mẫu negative mà stage classifier trước nó nhận dạng sai, tức là nó sẽ tập trung học từ các mẫu background khó hơn, do đó sự kết hợp các stage classifiers này lại sẽ giúp bộ phân loại có false alarm thấp. Với cấu trúc này,

những mẫu background dễ nhận diện sẽ bị loại ngay từ những stages đầu tiên, giúp đáp ứng tốt nhất đối với độ phức tạp gia tăng của các mẫu đưa vào, đồng thời giúp rút ngắn thời gian xử lý.

Thuật toán Cascade training:

1. Gọi:  
 $F$  là giá trị *false alarm* và  $d$  là độ chính xác của weak classifier ở mỗi stage  
 $F_{\text{target}}$ : Giá trị max false alarm.  
 $P, N$  là số lượng mẫu positive và negative.  
 $P_i, N_i$  là tập positive và negative cho bộ phân lớp ở tầng thứ  $i$ .  
 $F_i, D_i$ : Giá trị *false alarm* và độ chính xác của cascade trước khi đến tầng thứ  $i$ .
2. Khởi tạo  $i=0$ ;  $F_0=1.0$ ;  $D_0=1.0$
3. Lặp: while  $F_i > F_{\text{target}}$ 
  - $i = i+1$ ;
  - Huấn luyện bộ phân loại  $h_i$  từ tập  $P_i$  và  $N_i$  với detection rate  $d$  và max false alarm  $f$ . Thêm  $h_i$  vào cây phân lớp.
  - Dùng cây phân lớp hiện có để tính  $F_i$ : Duyệt qua  $N$  mẫu negative cho đến khi nào tìm đủ  $n$  mẫu mà cây phân lớp hiện có phân loại sai. ( $F_i = \frac{n}{N}$ )
  - $N := \emptyset$
  - Nếu  $F_i > F_{\text{target}}$   
 $N = \{ \text{số mẫu sai ở stage hiện tại phân loại sai} \}$   
 $P = \{ \text{số mẫu positive mà stage hiện tại phân loại đúng} \}$

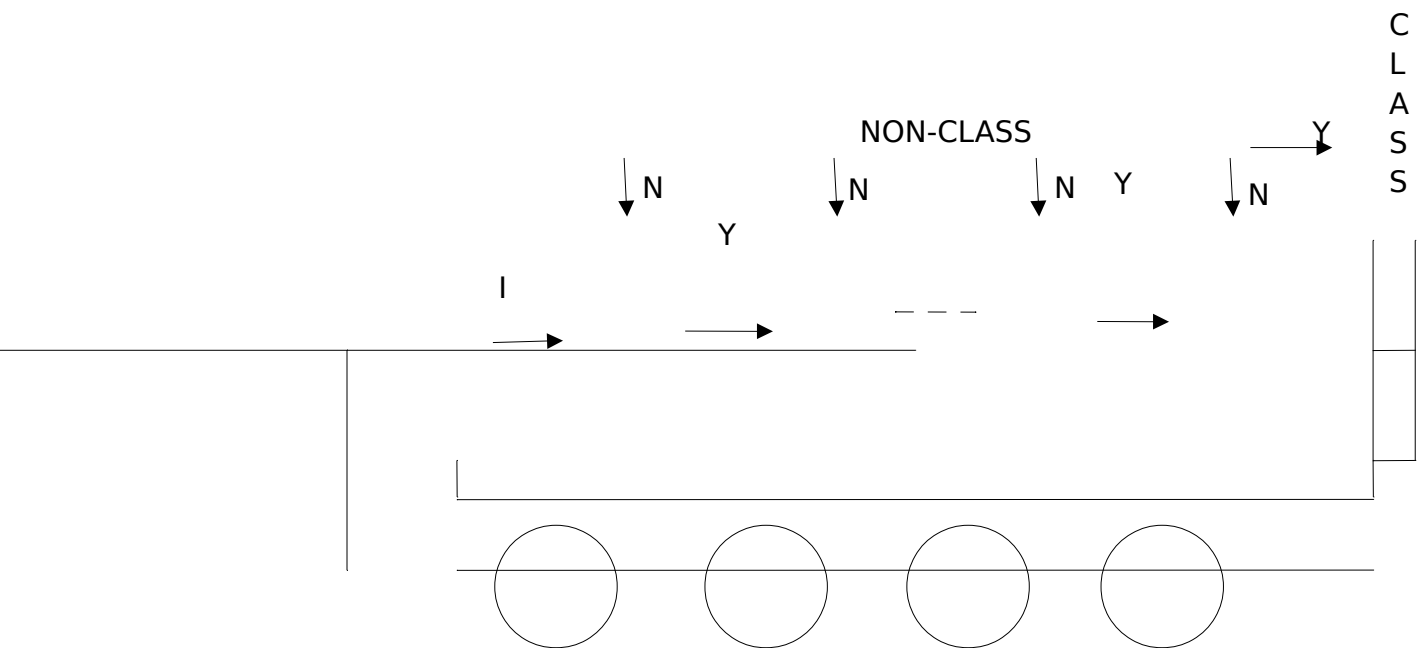
Minh họa thuật toán Cascade training:



Hình 18. Cascade of classifiers

### 3.1.5. Cascade of boosting classifiers

Cascade of boosting classifiers là mô hình cascade of classifiers với mỗi classifier được xây dựng bằng Adaboost sử dụng haar-like.

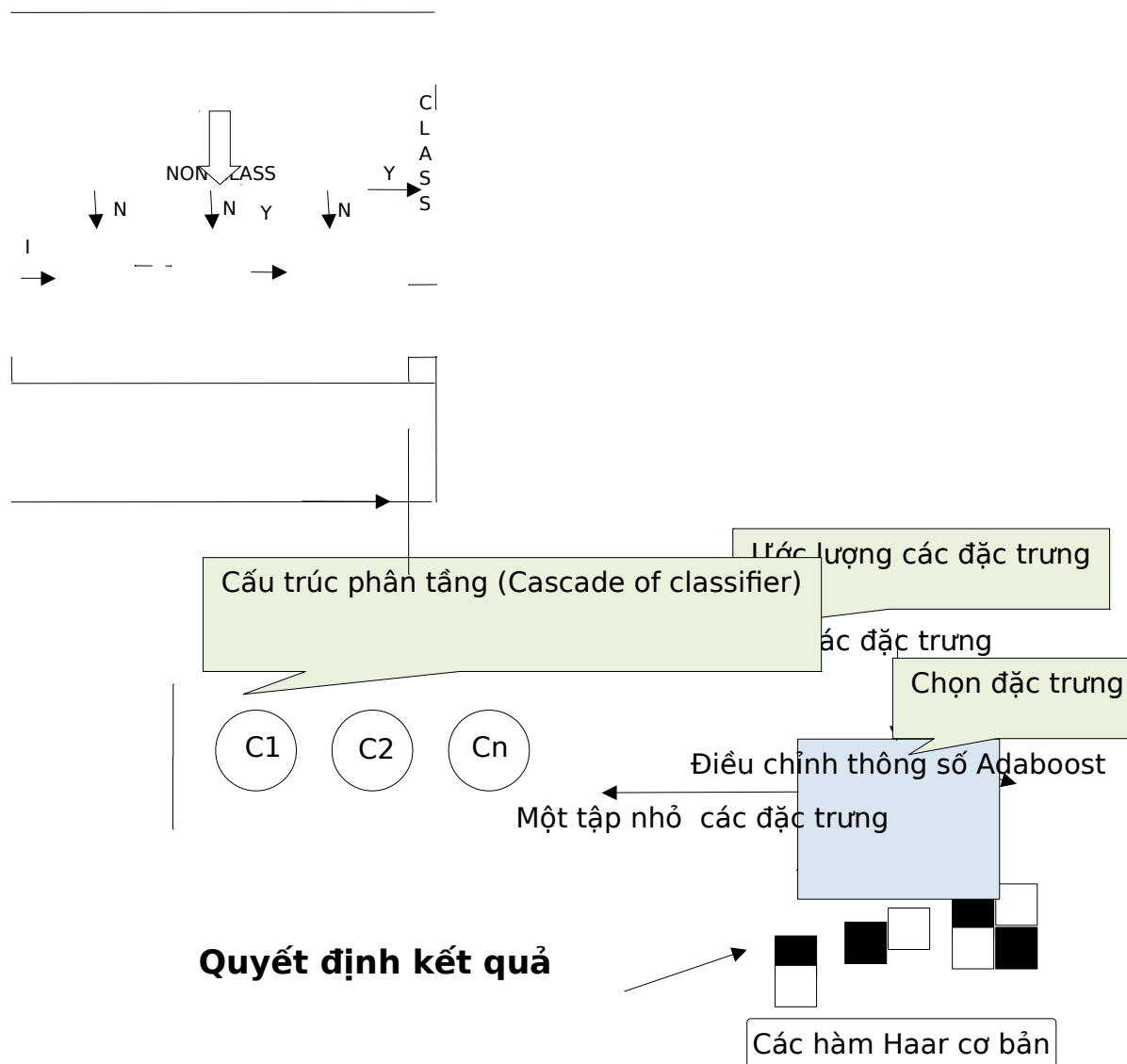


Hình 19. Cascade of boosting classifiers

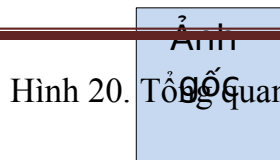
Viola và Jones đã sử dụng rất thành công cascade of boosting classifiers cho bài toán nhận dạng mặt người. Với tập huấn luyện gồm 4196 hình mặt người được đưa về ảnh grayscale kích thước 24x24 và 9500 hình background, hai ông đã xây dựng cấu trúc cascade tree gồm 38 stage với tổng cộng 6060 đặc trưng haar-like. Thực nghiệm đã cho thấy classifier ở stage đầu tiên sử dụng 2 đặc trưng và loại được khoảng 50% mẫu background (không phải mặt người) và có độ chính xác là  $d = 100\%$ . Classifier ở stage thứ 2 sử dụng 10 đặc trưng loại được 80% mẫu background với độ chính xác vẫn là 100%. Hệ thống này được so sánh với hệ thống của Rowley-Baluja-Kanade (sử dụng mạng neural), Schneiderman-Kanade (sử dụng phương pháp thống kê), và cho

thấy tỉ lệ nhận dạng là ngang nhau, trong khi hệ thống của Viola và Jones chạy nhanh hơn đến 15 lần so với hệ thống của Rowley-Baluja-Kanade và nhanh hơn 600 lần hệ thống của Schneiderman-Kanade.

Lý do mà cấu trúc cascade đạt tốc độ nhận dạng nhanh chính là nhờ nó sớm loại bỏ được các mẫu background đơn giản (thường có số lượng lớn hơn nhiều so với các mẫu chứa object – các mẫu chưa khuôn mặt cần tiến hành nhận dạng). Bên cạnh đó, hệ thống của Viola và Jones cũng đạt được độ chính xác khá cao nhờ vào thuật toán cascade training, các bộ nhận dạng được huấn luyện bằng AdaBoost với đặc trưng Haar-like mô tả tốt thông tin đối tượng, cùng với cách Integral Image tính nhanh các đặc trưng, không làm giảm tốc độ nhận dạng của hệ thống. Như vậy, mô hình Cascade of Boosted Classifiers thật sự là một cách tiếp cận tốt cả về tốc độ lẫn khả năng nhận dạng, rất phù hợp với bài nhận dạng mặt người.



Integral Image



Hình 20. Tổng quan về hệ thống phát hiện mặt người trong ảnh sử dụng Cascade of classifier

## 3.2. Nhận diện mặt người

Hai phương pháp tiếp cận với nhận dạng khuôn mặt hiện nay là: Nhận dạng dựa trên đặc trưng của các phần tử trên khuôn mặt (Feature Based Face Recognition), và nhận dạng dựa trên xét tổng thể toàn khuôn mặt (Appearance Based Face Recognition).

### 3.2.1. Nhận dạng dựa trên mối quan hệ giữa các phần tử (Feature Based)

Đây là phương pháp nhận dạng khuôn mặt dựa trên việc xác định các đặc trưng hình học của các chi tiết trên một khuôn mặt (như vị trí, diện tích, hình dạng của mắt, mũi, miệng, khuôn dạng của khuôn mặt,...), và mối quan hệ giữa chúng (ví dụ khoảng cách của hai mắt, khoảng cách của hai lông mày, khoảng cách từ mũi tới cằm,...).

#### ❖ Ưu điểm của phương pháp:

- Phương pháp này gần với cách mà con người sử dụng để nhận biết khuôn mặt.
- Với việc xác định đặc tính và các mối quan hệ, phương pháp này có thể cho kết quả tốt trong các trường hợp ảnh có nhiễu nhiều, ảnh bị nghiêng, bị xoay, ánh sáng thay đổi.

#### ❖ Nhược điểm của phương pháp:

- Việc xác định một lúc nhiều đặc tính và các mối quan hệ đòi hỏi các thuật toán phức tạp.
- Nếu ảnh có độ phân giải bé, các mối quan hệ giữa cùng một cặp đặc tính trên các khuôn mặt đều rất sát nhau, điều này dẫn đến kết quả nhận dạng không chính xác. Với ảnh có độ phân giải lớn hơn thì khối lượng tính toán sẽ rất lớn, đòi hỏi cao về cấu hình phần cứng.

### 3.2.2. Nhận dạng dựa trên xét toàn diện khuôn mặt (Appearance Based)

Nội dung chính của hướng tiếp cận này là xem mỗi ảnh có kích thước  $R \times C$  là một vec-tơ trong không gian  $R \times C$  chiều. Ta sẽ tìm một không gian mới có chiều nhỏ hơn, sao cho khi được biểu diễn trong không gian đó các đặc điểm chính của một khuôn



mặt vẫn không bị mất đi. Trong không gian mới đó, các ảnh của cùng một người sẽ tập hợp lại thành một nhóm gần nhau và cách xa các nhóm khác. Hai phương pháp thường được dùng hiện nay là PCA (Principle Components Analysis) và LDA (Linear Discriminant Analysis).

### 3.2.2.1. Nhận dạng bằng PCA

- ❖ PCA (Principle Components Analysis) là một thuật toán để tạo ra một ảnh mới từ ảnh ban đầu. Ảnh mới này có kích thước nhỏ hơn rất nhiều so với ảnh đầu vào và vẫn mang những đặc trưng cơ bản nhất của khuôn mặt cần nhận dạng. PCA không cần phải quan tâm đến việc phải tìm ra cụ thể các đặc điểm trên khuôn mặt (mắt, mũi, miệng...) và các mối quan hệ. Tất cả những chi tiết đó đều được thể hiện ở ảnh mới được tạo ra từ PCA.

Mục tiêu của PCA là ảnh mới được tạo ra có kích thước nhỏ nhất và chứa nhiều nhất các đặc trưng của ảnh khuôn mặt đầu vào. Về bản chất, PCA tìm ra một không gian mới theo hướng biến thiên mạnh nhất của một tập hợp các vec-tơ trong một không gian cho trước. Trong không gian mới, người ta hy vọng rằng việc phân loại sẽ mang lại kết quả tốt hơn so với không gian ban đầu.

#### ❖ Ưu điểm của phương pháp PCA:

- Tìm được các đặc tính tiêu biểu cho một khuôn mặt mà không cần phải xác định các thành phần và mối liên hệ giữa các thành phần trên khuôn mặt đó.
- Thuật toán thực hiện tốt với các ảnh đầu vào có độ phân giải cao, do PCA sẽ thu gọn ảnh đầu vào thành một ảnh mới có kích thước rất nhỏ.
- Khối lượng tính toán không nhiều.
- PCA có tính mở và có thể kết hợp được với các phương pháp khác (như mạng nơ-rôn, Support Vector Machine – SVM,...) để mang lại hiệu quả nhận dạng cao hơn.

#### ❖ Nhược điểm của PCA:

- PCA phân loại theo chiều phân bố lớn nhất của tập vec-tơ. Tuy nhiên, không phải bao giờ chiều phân bố lớn nhất lại mang lại hiệu quả cao nhất cho nhận dạng. đây là nhược điểm cơ bản của PCA.

- PCA rất nhạy với nhiễu, vì vậy nhiệm vụ tiền xử lý (lọc nhiễu, chuẩn hóa, ...) đòi hỏi phải thực hiện rất kỹ

### 3.2.2.2. Nhận dạng bằng LDA

Tương tự như PCA, LDA (Linear Discriminants Analysis) cũng là một phương pháp nhận dạng theo hướng tiếp cận Appearance Based.

Mục đích của LDA là tìm ra một không gian mới để tách rời các nhóm ảnh với nhau. Không gian đó được tạo bởi các vec-tơ không nhất thiết phải trực giao từng đôi một.

Trong không gian tạo ra từ LDA, các ảnh trong cùng một nhóm sẽ gần nhau, và các nhóm khác nhau sẽ cách xa nhau. LDA khắc phục được nhược điểm của PCA, đó là “không phải hướng biến thiên mạnh nhất luôn mang thông tin nhận dạng cao nhất”.

#### ❖ Ưu điểm của phương pháp:

- LDA mang hầu hết các ưu điểm của phương pháp PCA, đồng thời khắc phục được nhược điểm cơ bản của PCA.

#### ❖ Nhược điểm của phương pháp:

- LDA đòi hỏi một khối lượng tính toán rất lớn. Hiện nay vẫn chưa có cách tính nhanh phương pháp LDA.
- LDA cũng nhạy với nhiễu giống như PCA, điều này đòi hỏi nhiệm vụ tiền xử lý ảnh phải được thực hiện kỹ lưỡng.

### 3.2.3. Nội dung phương pháp PCA

#### 3.2.3.1. Xây dựng phương pháp PCA

Trong nhận dạng mặt người, khi sử dụng PCA, ta coi một ảnh xám  $R \times C$  pixels là một vec-tơ trong không gian  $N = R \times C$  chiều.

Không gian mới được tạo ra bởi PCA được cấu thành từ  $K$  vec-tơ đơn vị có chiều là  $N$ ,  $N \ll R \times C$ . Mỗi vec-tơ được gọi là một Eigenface.

Phép biến đổi:

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_N \end{bmatrix} \longrightarrow \mathbf{W} = \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_K \end{bmatrix} \text{ với } K \ll N$$

Theo công thức:  $\mathbf{W} = \mathbf{T} \cdot \mathbf{A}$

Với  $\mathbf{T}$  là ma trận chuyển đổi.  $\mathbf{T}$  có kích thước là  $K \times N$

- Gọi  $M$  là số ảnh đầu vào, mỗi ảnh đã được chuyển thành một vec-tơ  $N$  chiều. Ta có tập hợp đầu vào  $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M\}$  ( $\mathbf{x}_i \in \mathbb{R}^N$ ).

- Trung bình của các vec-tơ đầu vào (tâm của tập hợp đầu vào):

$$\bar{\mathbf{x}} = \frac{1}{M} \sum_{i=1}^M \mathbf{x}_i$$

- Sai lệch so với tâm:

$$\Phi_i = \mathbf{x}_i - \bar{\mathbf{x}}$$

- Gọi  $\mathbf{A} = [\Phi_1, \Phi_2, \dots, \Phi_N]$ , ta có covariance matrix của  $\mathbf{A}$ :

$$\mathbf{C} = \frac{1}{M} \sum_{i=1}^M \Phi_i \Phi_i^T = \mathbf{A} \cdot \mathbf{A}^T$$

- Gọi các giá trị riêng của  $\mathbf{C}$  là  $\lambda_1, \lambda_2, \dots, \lambda_N$  sắp xếp theo thứ tự giảm dần, tương

ứng với  $N$  vec-tơ riêng  $u_1, u_2, \dots, u_N$ . Các vec-tơ riêng này trực giao từng đôi một. Mỗi vec-tơ riêng  $u_i$  được gọi là một eigenface. Tập hợp vec-tơ ban đầu được biểu diễn trong không gian tạo bởi  $N$  eigenface theo mô tả sau:

$$\mathbf{x} - \bar{\mathbf{x}} = w_1 \mathbf{u}_1 + w_2 \mathbf{u}_2 + \dots + w_N \mathbf{u}_N = \sum_{i=1}^N w_i \mathbf{u}_i$$

➤ Chọn lấy  $K$  vec-tơ riêng  $u$  tương ứng với  $K$  giá trị riêng  $\lambda$  lớn nhất. Ta có:

$$\hat{\mathbf{x}} - \bar{\mathbf{x}} = w_1 \mathbf{u}_1 + w_2 \mathbf{u}_2 + \dots + w_K \mathbf{u}_K = \sum_{i=1}^K w_i \mathbf{u}_i \text{ với } K \ll N$$

Véc-tơ các hệ số khai triển  $[w_1, w_2, \dots, w_K]$  chính là biểu diễn mới của ảnh được tạo ra trong không gian PCA. Véc-tơ  $[w_1, w_2, \dots, w_K]$  được tính theo công thức sau:

$$\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_K \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \vdots \\ \mathbf{u}_K^T \end{bmatrix} (\mathbf{x} - \bar{\mathbf{x}}) = \mathbf{U}^T \cdot (\mathbf{x} - \bar{\mathbf{x}})$$

Vấn đề cần giải quyết ở đây là ma trận covariance  $C = A \cdot A^T$  có kích thước  $N^2$ . Với  $N = 200 \times 200 = 40000$ , khối lượng tính toán sẽ rất lớn. Do đó, để tính được các eigenface mà không cần phải tính cả ma trận  $C$ , người ta đưa ra phương pháp tính nhanh dựa vào tính vec-tơ riêng và giá trị riêng của ma trận  $L = A^T \cdot A$  có kích thước  $M \times M$ , với  $M$  là số ảnh đầu vào. Ta có thể chứng minh như sau: gọi  $v_i, \mu_i$  lần lượt là vec-tơ riêng và giá trị riêng của ma trận  $L$ :

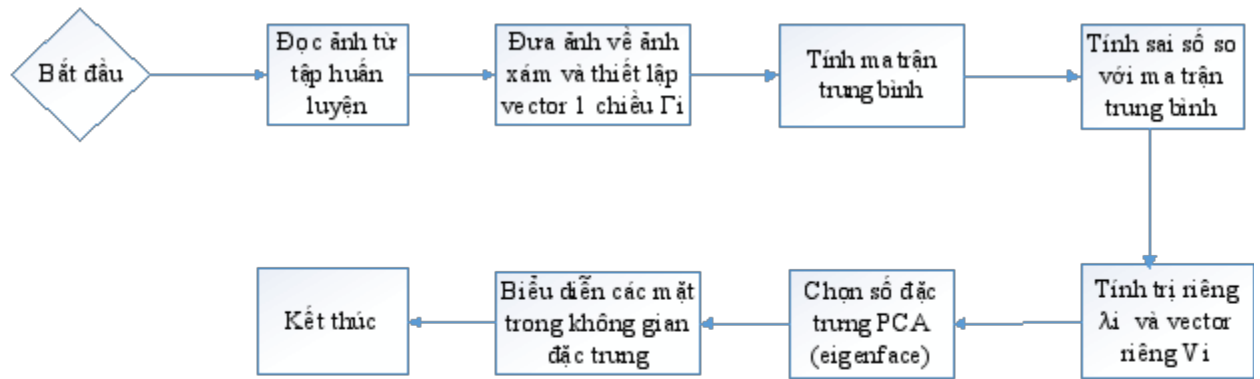
$$A^T \cdot A \cdot v_i = \mu_i \cdot v_i$$

Nhân cả 2 vế với  $A$  ta có:

$$A \cdot A^T \cdot A \cdot v_i = \mu_i \cdot A \cdot v_i$$

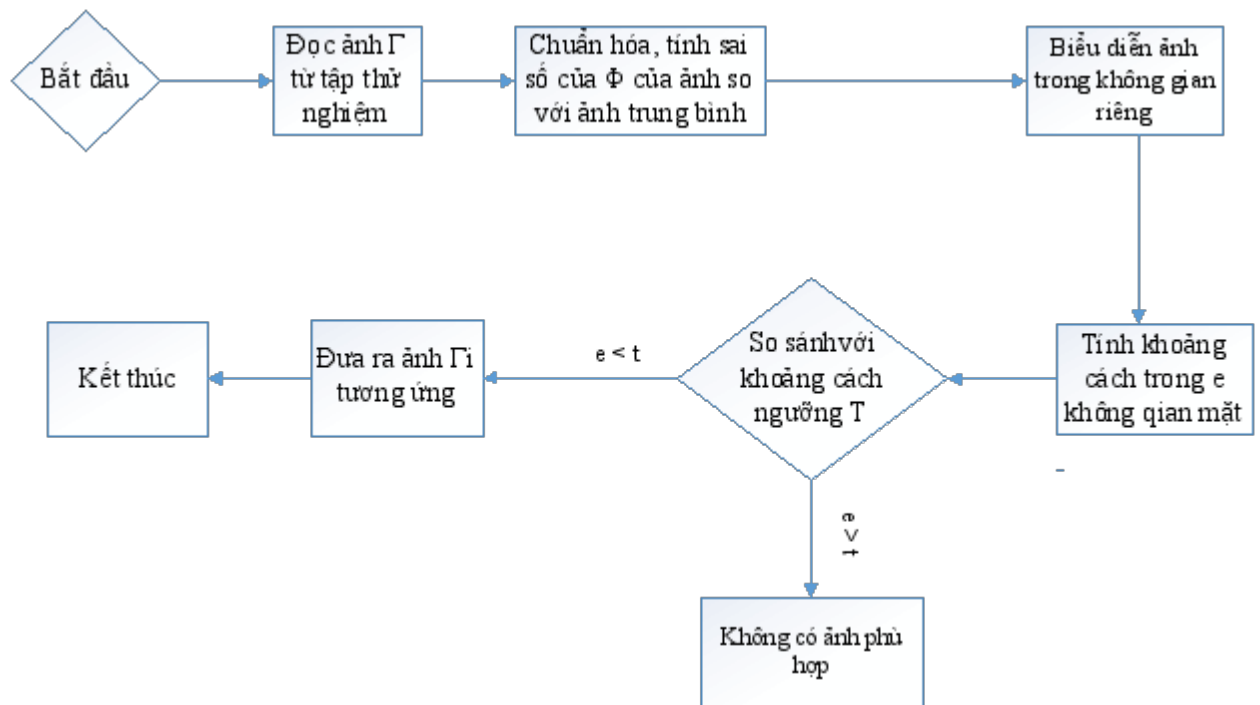
Ta thấy  $A \cdot v_i$  chính là vec-tơ riêng của  $C = A \cdot A^T$ , ứng với giá trị riêng  $\mu_i$

## 3.2.3.2. Lưu đồ thuật toán PCA



Hình 21. Lưu đồ thuật toán tính đặc trưng mặt trong tập huấn luyện

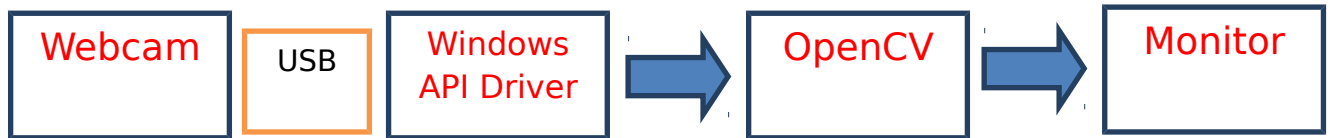
((original faces) là hình ảnh khuôn mặt ban đầu)



Hình 22. Lưu đồ thuật toán tính nhận diện hình ảnh khuôn mặt

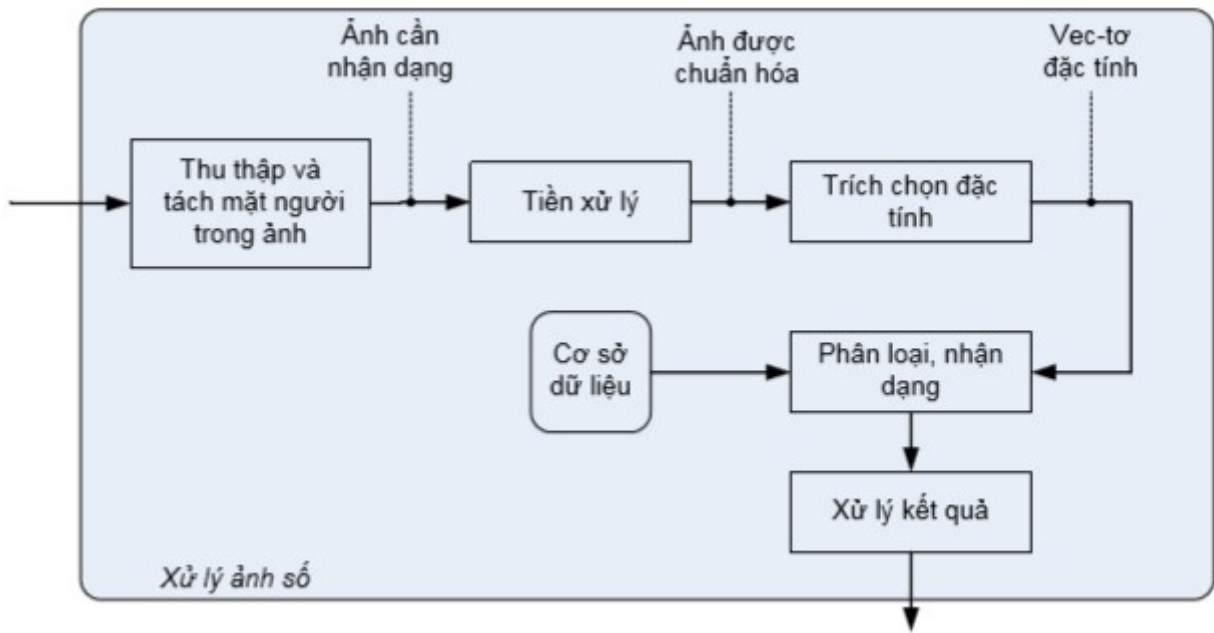
## 3.3. Sơ đồ khối của hệ thống

### 3.3.1. Sơ đồ khối phần cứng



Hình 23. Sơ đồ khối cho phần cứng

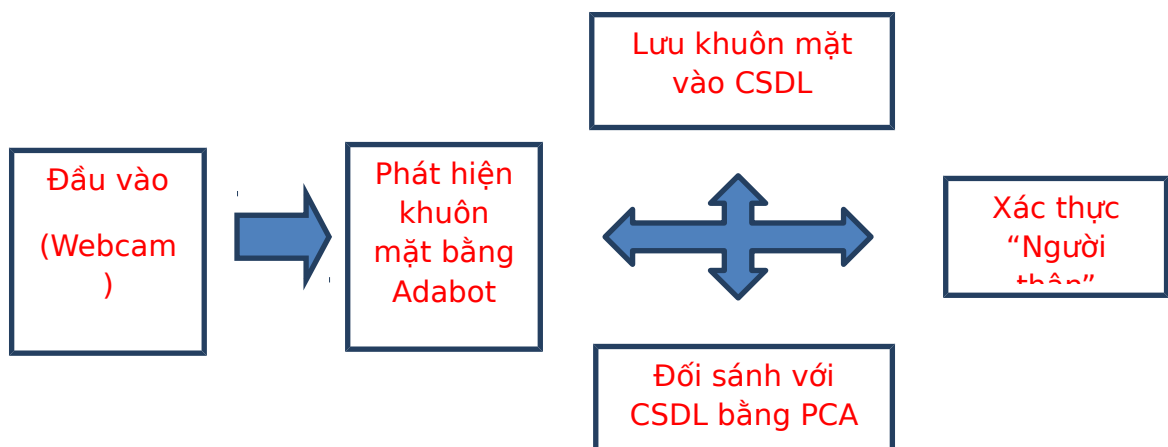
### 3.3.2. Sơ đồ khối phần mềm



Hình 24. Sơ đồ khối cho ứng dụng

### 3.3.3. Thiết kế hệ thống cho chương trình nhận diện khuôn mặt

Chương trình được chia thành thành các phần như sau:



Hình 25. Các thành phần chương trình nhận diện mặt người

➤ *Xử lý đầu vào:*

Chương trình nhận đầu vào là webcam, tuy nhiên, việc phát hiện khuôn mặt được thực hiện trên các bức ảnh, do đó, ta phải chuyển thành các ảnh tĩnh và xử lý trên từng ảnh tĩnh. Sau khi đã có ảnh đầu vào rồi thì sẽ chuyển ảnh cho giai đoạn xử lý tiếp theo, đó là phát hiện các khuôn mặt có trong ảnh.

➤ *Phát hiện khuôn mặt:*

Phần này xử lý chính nghiệp vụ của hệ thống. Sau khi có ảnh truyền vào, hệ thống sẽ thực hiện chức năng phát hiện khuôn mặt có trong ảnh.

Việc phát hiện khuôn mặt được thực hiện nhanh bằng thuật toán adaboost thông qua hàm `cvHaarDetectObjects()` của OpenCV. Hàm này thực hiện việc phát hiện đối tượng dựa trên các đặc trưng haar-like, cụ thể là nhờ vào một bộ Cascade được truyền vào cho hàm. Bộ Cascade được xây dựng theo dạng cây (tree-node) và đã được huấn luyện từ trước.

Việc huấn luyện bộ Cascade có thể thực hiện từ những dữ liệu thu thập được để phục vụ cho quá trình nhận dạng. Ví dụ, muốn nhận dạng một người A, ta thu thập các ảnh khuôn mặt của người A với nhiều tư thế, góc chụp và điều kiện chụp khác nhau, sau đấy cho bộ nhận dạng học theo thuật toán Cascade training.

➤ *Lưu khuôn mặt vào cơ sở dữ liệu:*

Khuôn mặt sau khi được phát hiện sẽ được tách ra khỏi bức ảnh và lưu dưới dạng ảnh bitmap với phần mở rộng là \*.jpg.

Các khuôn mặt sẽ được tách ra và lưu lại thành các file ảnh riêng rẽ, và được đánh số thứ tự theo tên khuôn mặt.

➤ *Đối sánh với cơ sở dữ liệu:*

Sau khi quá trình lưu trữ hình ảnh vào cơ sở dữ liệu hoàn tất, hệ thống tiếp tục quay lại giai đoạn xử lý đầu vào, phát hiện khuôn mặt. Sử dụng thuật toán PCA để đổi chiều và so sánh giữa khuôn mặt hiện tại và những khuôn mặt đang có trong cơ sở dữ liệu. Sau đó PCA sẽ tìm ra được tỉ lệ giống nhau giữa hai bên.

### ➤ *Xác thực người thân:*

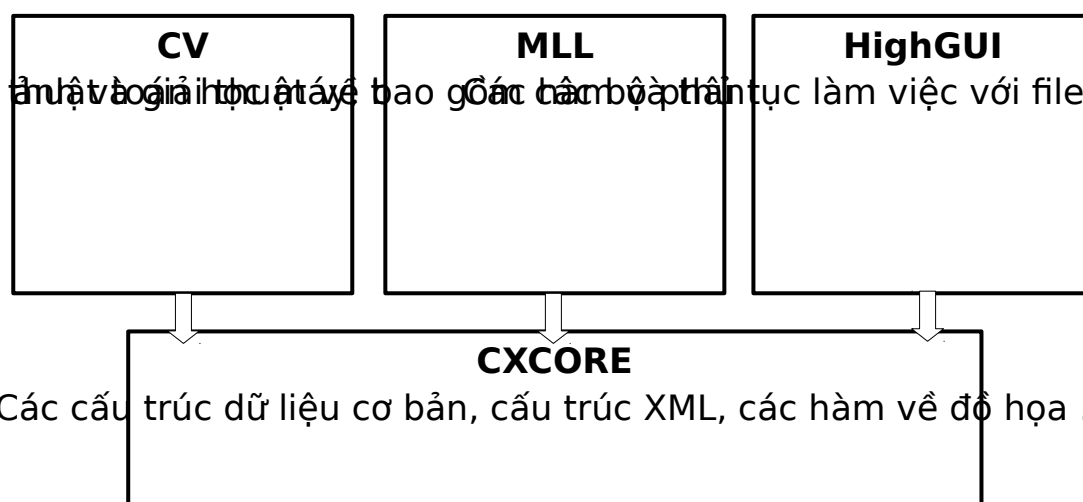
Nếu tỉ lệ giống nhau mà thuật toán PCA tìm được nằm trong một ngưỡng quy định cho phép, chúng ta xác thực được người đó chính là người đã từng được lưu trong cơ sở dữ liệu, và ngược lại.

## CHƯƠNG 4. XÂY DỰNG THỰC NGHIỆM

### 4.1. Sơ lược về OpenCV

OpenCV là thư viện mã nguồn mở của intel về thị giác máy tính. Nó cung cấp một bộ mã nguồn bao gồm hàng trăm hàm, lớp dựa trên các thuật toán về xử lý ảnh cũng như Computer vision dùng ngôn ngữ C/C++. Open CV thể hiện sự đa dạng của trí tuệ nhân tạo. Được ứng dụng nhiều trong các bài toán nhận dạng mặt, dò tìm mặt, phát hiện mặt, lọc Kalman, ...

Cấu trúc tổng quan của OpenCV bao gồm 5 phần chính. 4 trong 5 phần đó được chỉ ra trong hình vẽ dưới.



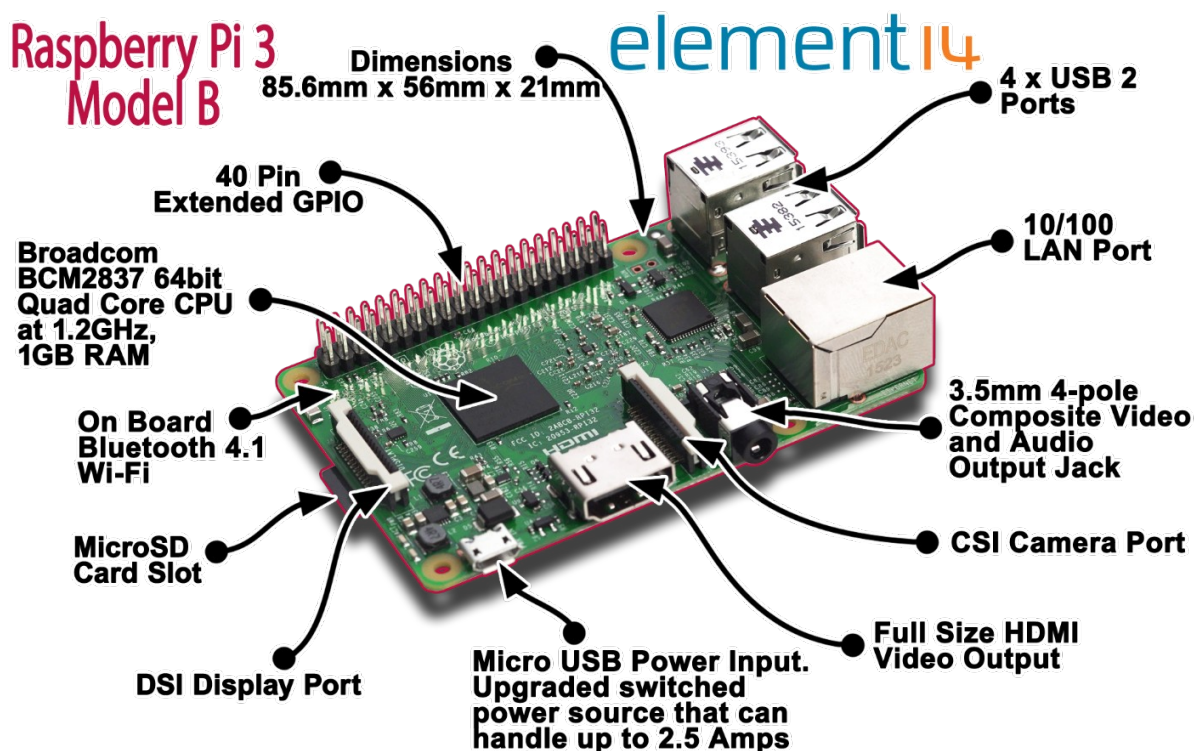
Hình 26. Cấu trúc cơ bản của OpenCV

Phần CV bao gồm các thư viện cơ bản về xử lý ảnh và các giải thuật về thị giác máy tính. ML là bộ thư viện về các thuật toán học máy, bao gồm rất nhiều bộ phân cụm và phân loại thống kê. HighGUI chứa đựng những thủ tục vào ra, các chức năng về lưu trữ cũng như đọc các file ảnh và video. Phần thứ 4, Cxcore chứa đựng các cấu trúc dữ liệu cơ bản ( ví dụ như cấu trúc XML, các cây dữ liệu ...). Phần cuối cùng là CvAux,



phần này bao gồm các thư viện cho việc phát hiện, theo dõi và nhận dạng đối tượng (khuôn mặt, mắt ...).

### 4.2. Sơ lược về phần cứng Kit Raspberry Pi 3



Hình 27. Raspberry Pi 3

Thông số kỹ thuật chi tiết:

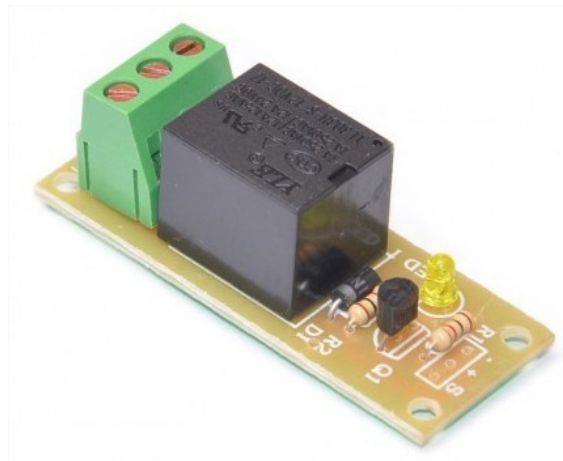
- 1.2 GHz 64-bit quad-core ARM Cortex-A53 CPU (BCM2837)
- 1 GB RAM (LPDDR2 SDRAM)
- On-board Wireless LAN – 2.4 GHz 802.11 b/g/n (BCM43438)
- On-board Bluetooth 4.1 + HS Low-energy (BLE) (BCM43438)
- 4 x USB 2.0 ports
- 10/100 Ethernet
- 40 GPIO pins
- Full size HDMI 1.3a port
- Combined 3.5mm analog audio and composite video jack
- Camera interface (CSI)
- Display interface (DSI)
- MicroSD slot
- VideoCore IV multimedia/3D graphics core @ 400MHz/300MHz

## 4.3. Sơ lược về bộ điều khiển mở cửa Rơ-le

Giới thiệu:

- Rơ-le là một loại linh kiện điện tử thụ động rất hay gặp trong các ứng dụng thực tế. Khi bạn gặp các vấn đề liên quan đến công suất và cần sự ổn định cao, ngoài ra có thể dễ dàng bảo trì, thì rơ-le chính là cái bạn cần tìm
- Rơ-le là một **công tắc** (khóa K). Nhưng khác với công tắc ở một chỗ cơ bản, rơ-le được kích hoạt bằng điện thay vì dùng tay người. Chính vì lẽ đó, rơ-le được dùng làm công tắc điện tử! Vì rơ-le là một công tắc nên nó có 2 trạng thái: **đóng** và **mở**.

Trên thị trường chúng ta có 2 loại module rơ-le: **module rơ-le đóng ở mức thấp** (nối cực âm vào chân tín hiệu rơ-le sẽ đóng), **module rơ-le đóng ở mức cao** (nối cực dương vào chân tín hiệu rơ-le sẽ đóng)



Module rơ-le kích ở mức cao

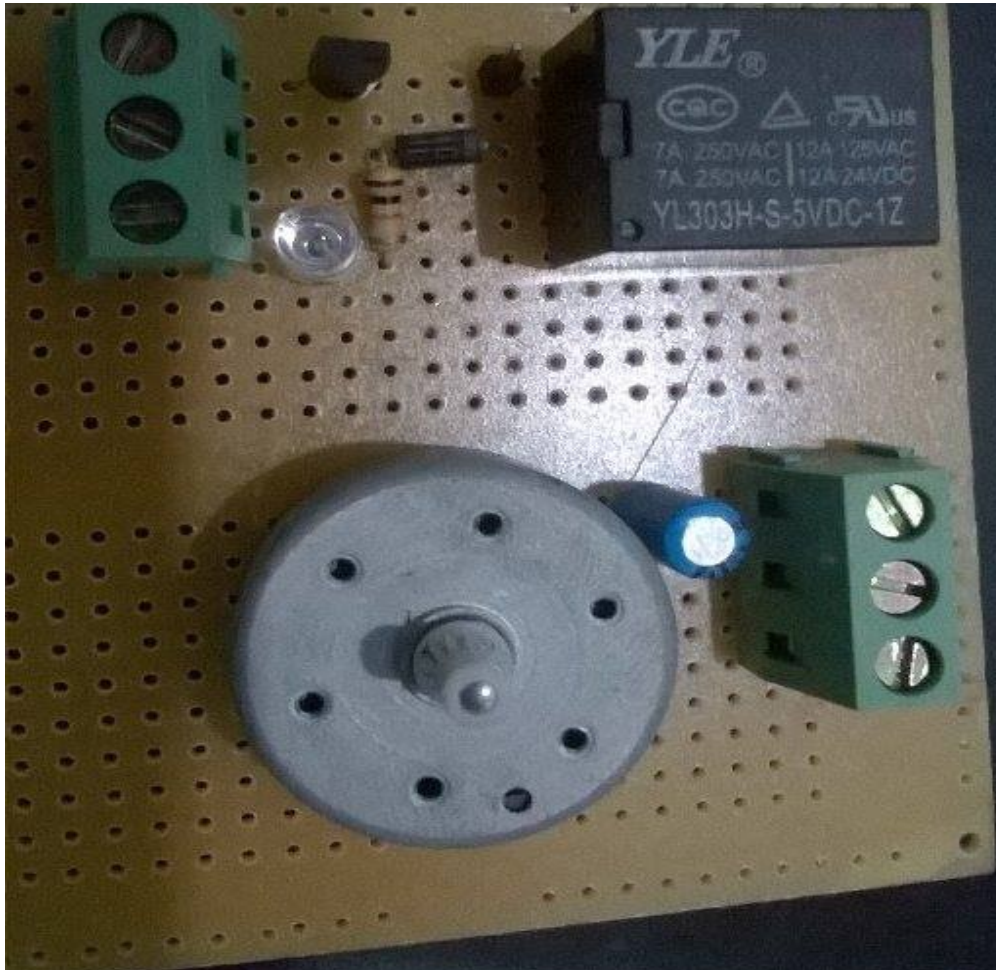


Module rơ-le kích ở mức thấp

Rơ-le bình thường gồm có 5 chân. Trong đó có 3 chân để kích, 2 chân còn lại nối với đồ dùng điện

- 3 chân dùng để kích
  - +: cấp hiệu điện thế kích tối ưu vào chân này.
  - - : nối với cực âm
  - S: chân tín hiệu, tùy vào loại module rơ-le mà nó sẽ làm nhiệm vụ kích rơ-le
- 1 trong 2 chân còn lại sẽ nối với thiết bị điện

Sản phẩm thực tế của nhóm



Hình 28. Bộ điều khiển Rơ-le

### 4.4. Sơ lược Webcam Logitech C270



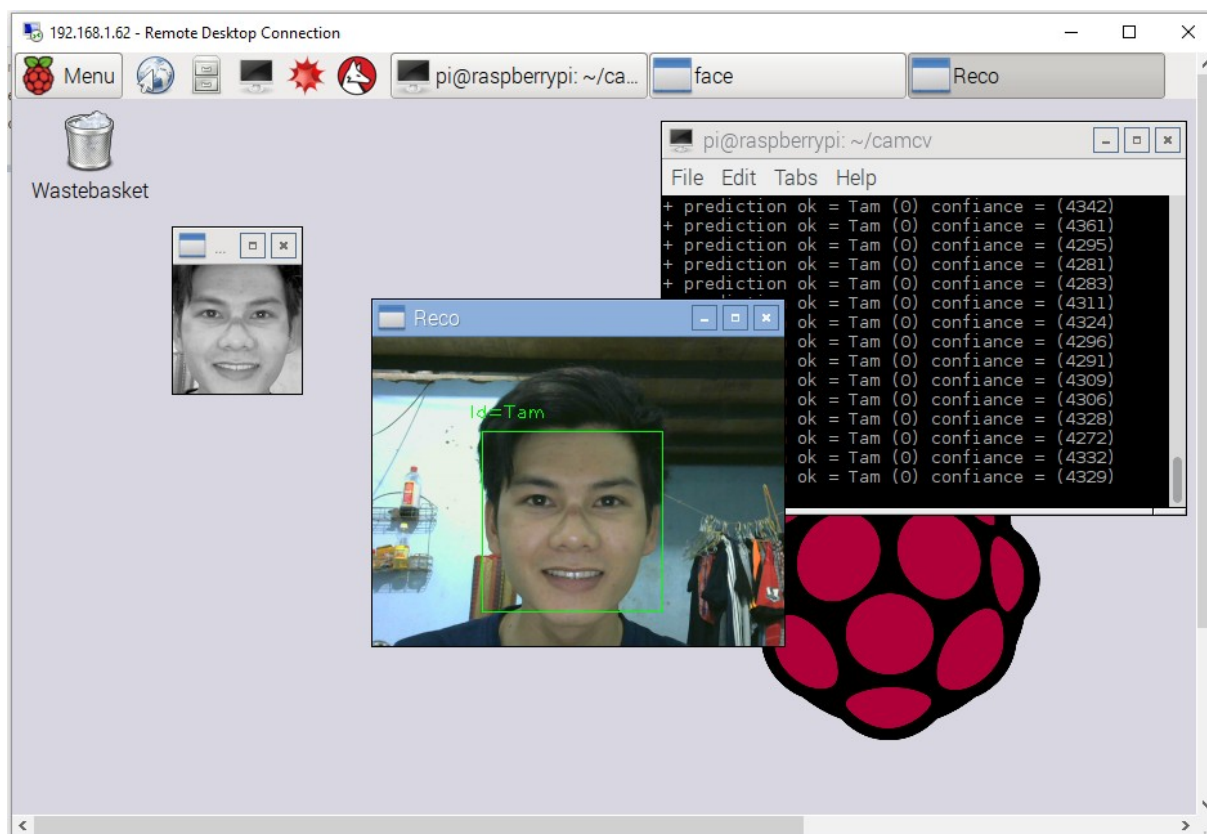
Hình 29. Webcam Logitech C270

Thông tin kỹ thuật:

- Chụp ảnh ở chế độ 3Megapixel
- Điện thoại hình ảnh ở chế độ HD 720p
- Ứng dụng công nghệ Right Light - tự động điều chỉnh để lấy được ánh sáng tối ưu và cho ảnh tốt ngay cả trong điều kiện ánh sáng mập mờ.
- Tích hợp micro với công nghệ RightSound™
- Thương hiệu Logitech

### 4.5. Kết quả thực nghiệm

Sau khi phát hiện được khuôn mặt, chúng ta thực hiện lệnh train cho khuôn mặt này, xử lý cắt ảnh, chỉnh size, đưa về ảnh xám và các thông số ảnh. Lưu thông tin vào cơ sở dữ liệu.

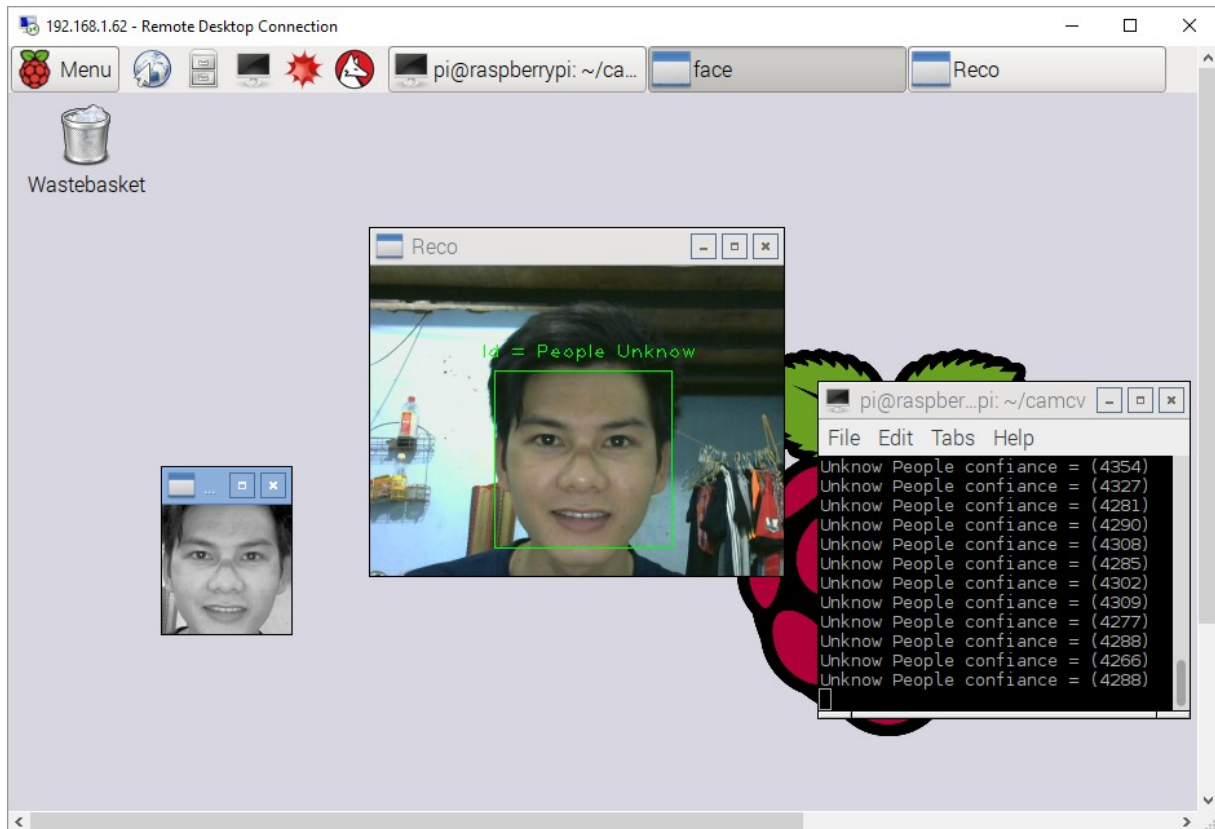


Hình 30. Nhận diện là người thân

Khi phát hiện được một khuôn mặt mới, tiếp tục xử lý và phân tích ảnh. Sau đó đối sánh khuôn mặt hiện tại với những khuôn mặt trong cơ sở dữ liệu, nếu giống thì đó là người thân. Trong trường hợp này, người thân chính là tên chứa khuôn mặt đã được train trước đó. Ngược lại, là người lạ.

Nếu chương trình nhận diện đúng là người thân, robot sẽ cho phép mở cửa, đồng thời phát ra giọng nói đã được lập trình trước, tạo được hiệu ứng thú vị và chuyên nghiệp. Sau một thời gian quy định chương trình sẽ cho phép đóng cửa lại, đồng nghĩa với việc người thân đã đi vào nhà và cửa sẽ tự đóng.





Hình 31. Nhận diện không phải là người thân

Nếu không phải người thân, cửa sẽ không được mở, đồng thời phát ra giọng nói “You can not open the door”.

### CHƯƠNG 5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

#### 5.1. Kết luận

Trên cơ sở nghiên cứu về bài toán nhận diện mặt người. Đặc biệt là các thuật toán Adaboost, các đặc trưng *haar-like* và mô hình *Cascade of Classifier* trong phát hiện mặt người và thuật toán PCA trong nhận diện mặt người. Nhóm đồ án đã áp dụng thành công và xây dựng được chương trình nhận diện người thân. Bên cạnh đó, nhóm đã tìm hiểu về thư viện mã nguồn mở OpenCV, hệ quản trị cơ sở dữ liệu Oracle, qua đó biết cách sử dụng các hàm trong thư viện OpenCV và quản lý cơ sở dữ liệu bằng hệ quản trị cơ sở dữ liệu Oracle.

Về sản phẩm, sau khi thử nghiệm trực tiếp trên kit Raspberry Pi 3. Chương trình đạt kết quả tốt, thời gian phát hiện nhanh đối với các ảnh mặt người chụp thẳng, chất lượng ảnh tốt, tuy nhiên đối với các ảnh chụp nghiêng hay ảnh có màu sắc quá tối thì chương trình gần như không thể phát hiện được khuôn mặt trong bức ảnh đấy.

Tuy nhiên, chương trình còn tùy theo chất lượng của loại webcam được sử dụng, chương trình sẽ không thể phát hiện mặt người trong điều kiện độ sáng quá yếu và góc quay nghiêng quá  $20^\circ$ .

#### 5.2. Hướng phát triển

Công việc kế tiếp của nhóm sẽ là thực hiện đánh giá thời gian thực hiện của các thuật toán nhận dạng trên các tập dữ liệu khác nhau. Đồng thời có thể kết hợp với các hệ thống điều khiển Robot để nhận dạng đối tượng như: Nhận dạng chủ nhân của Robot, nhận dạng và định danh tội phạm, hệ thống bảo mật bằng nhận diện khuôn mặt trong những ngôi nhà thông minh.

- Mở rộng số lượng người cần nhận dạng.
- Giải quyết mở rộng các khả năng xử lý của chương trình với các trường hợp thay đổi tư thế, trạng thái cảm xúc của đối tượng và giảm thiểu ảnh hưởng của phức tạp.
- Thử nghiệm, phối hợp các mô hình nhận dạng khác nhau (mạng nơ-rôn, SVM, ...) và các phương pháp nhận dạng ảnh khác (LDA, ICA,...) để mang lại độ tin cậy cao hơn cho hệ thống.

- Quy định ánh sáng tại những nơi có sử dụng hệ thống để giảm thiểu sai số nhận dạng do cường độ sáng thay đổi.
- “Nhúng” hệ thống xuống các hệ vi xử lý để ta có được một hệ thống chuyên dụng.
- Giải quyết vấn đề về nhận diện khuôn mặt đối với những dân tộc có màu da khác nhau, điển hình như nhận diện người châu phi có màu da đen.



## TÀI LIỆU THAM KHẢO

Đồ án tìm hiểu đặc trưng lõm 3D và bài toán phát hiện mặt người trong ảnh - Nguyễn Thị Thơm(2013)

Learning OpenCV - Gary Bradski and Adrian Kaehler

Ứng dụng xử lý ảnh trong thực tế với thư viện OpenCV C/C++ - Nguyễn Văn Long

Mastering OpenCV with Practical Computer Vision Projects – Sherin Emami:

<http://www.shervinemami.info/faceRecognition.html>

Face Recognition with OpenCV:

[http://docs.opencv.org/2.4/modules/contrib/doc/facerec/facerec\\_tutorial.html](http://docs.opencv.org/2.4/modules/contrib/doc/facerec/facerec_tutorial.html)

Face Detection using Haar Cascades:

[http://docs.opencv.org/master/d7/d8b/tutorial\\_py\\_face\\_detection.html#gsc.tab=0](http://docs.opencv.org/master/d7/d8b/tutorial_py_face_detection.html#gsc.tab=0)

Thông tin phần cứng kit Raspberry Pi 3

<http://hshop.vn/products/raspberry-pi-3-made-in-uk>

Cách hoạt động của Rơ-le

<http://arduino.vn/bai-viet/302-module-relay-cach-su-dung-ro-le-va-nhung-ung-dung-hay-cua-no>

## Hết