

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC DÂN LẬP HẢI PHÒNG**

-----o0o-----



ISO 9001:2015

**XÂY DỰNG ỨNG DỤNG
PHÁT HIỆN KHUÔN MẶT TRONG ẢNH SỬ
DỤNG OPENCV**

ĐỒ ÁN TỐT NGHIỆP ĐẠI HỌC HỆ CHÍNH QUY

Ngành: Công nghệ Thông tin

HẢI PHÒNG - 2019

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC DÂN LẬP HẢI PHÒNG**

-----o0o-----

**XÂY DỰNG ỨNG DỤNG
PHÁT HIỆN KHUÔN MẶT TRONG ẢNH SỬ
DỤNG OPENCV**

ĐỒ ÁN TỐT NGHIỆP ĐẠI HỌC HỆ CHÍNH QUY

Ngành: Công nghệ Thông tin

Sinh viên thực hiện : Cao Tiến Đạt
Mã sinh viên : 1512111009
Giáo viên hướng dẫn : TS. Ngô Trường Giang.

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC DÂN LẬP HẢI PHÒNG
-----o0o-----

NHIỆM VỤ THIẾT KẾ TỐT NGHIỆP

Sinh viên: Cao Tiến Đạt

Mã sinh viên: 1512111009

Lớp: CT1901C

Ngành: Công nghệ Thông tin

Tên đề tài: “Xây dựng ứng dụng phát hiện khuôn mặt trong ảnh sử dụng OpenCV”

NHIỆM VỤ ĐỀ TÀI

1. Nội dung và các yêu cầu cần giải quyết trong nhiệm vụ đề tài tốt nghiệp

a. Nội dung:

- Tìm hiểu tổng quan về phát hiện khuôn mặt trong ảnh
- Tìm hiểu phương pháp phát hiện khuôn mặt dựa trên đặc trưng Haarlike
- Tìm hiểu một số hàm cơ bản trong thư viện OpenCV sử dụng cho phát hiện khuôn mặt trong ảnh.

b. Các yêu cầu cần giải quyết

- Trình bày tổng quan về phát hiện khuôn mặt trong ảnh
- Hiểu và trình bày phương pháp phát hiện khuôn mặt dựa trên đặc trưng.
- Xây dựng ứng dụng phát hiện khuôn mặt trong ảnh sử dụng OpenCV.

2. Các số liệu cần thiết để thiết kế, tính toán

3. Địa điểm thực tập

CÁN BỘ HƯỚNG DẪN ĐỀ TÀI TỐT NGHIỆP

Người hướng dẫn thứ nhất:

Họ và tên: Ngô Trường Giang

Học hàm, học vị: Tiến sĩ.

Cơ quan công tác: Khoa Công nghệ Thông tin

Nội dung hướng dẫn:

- Tìm hiểu tổng quan về phát hiện khuôn mặt trong ảnh
- Tìm hiểu phương pháp phát hiện khuôn mặt dựa trên đặc trưng
- Tìm hiểu một số hàm cơ bản trong thư viện OpenCV sử dụng cho phát hiện khuôn mặt trong ảnh.

Người hướng dẫn thứ hai:

Họ và tên:

Học hàm, học vị.....

Cơ quan công tác:

Nội dung hướng dẫn:

.....
.....

Đề tài tốt nghiệp được giao ngày 01 tháng 7 năm 2019

Yêu cầu phải hoàn thành trước ngày 21 tháng 9 năm 2019

Đã nhận nhiệm vụ: Đ.T.T.N
Sinh viên

Đã nhận nhiệm vụ: Đ.T.T.N
Cán bộ hướng dẫn Đ.T.T.N

Cao Tiến Đạt

Ngô Trường Giang

Hải Phòng, ngàytháng.....năm 2019

HIỆU TRƯỞNG

GS.TS.NGUT Trần Hữu Nghị

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập – Tự do – Hạnh phúc

PHIẾU NHẬN XÉT CỦA CÁN BỘ HƯỚNG DẪN TỐT NGHIỆP

Họ và tên: Ngô Trường Giang

Cơ quan công tác: Khoa Công nghệ Thông tin

Họ tên sinh viên: Cao Tiến Đạt Ngành: Công nghệ Thông tin Nội dung hướng dẫn:

- Tìm hiểu tổng quan về phát hiện khuôn mặt trong ảnh
- Tìm hiểu phương pháp phát hiện khuôn mặt dựa trên đặc trưng
- Tìm hiểu một số hàm cơ bản trong thư viện OpenCV sử dụng cho phát hiện khuôn mặt trong ảnh.

1. Tinh thần thái độ của sinh viên trong quá trình làm đề tài tốt nghiệp:

- Sinh viên chủ động tìm đọc các tài liệu liên quan tới đề tài
- Chấp hành nghiêm túc kế hoạch, tiến độ đề ra.

2. Đánh giá chất lượng của đồ án (so với nội dung yêu cầu đã đề ra trong nhiệm vụ đề tài tốt nghiệp trên các mặt lý luận, thực tiễn, tính toán số liệu..):

- Về mặt lý thuyết: Đồ án trình bày các vấn đề cơ bản về Phát hiện khuôn mặt trong ảnh, đặc trưng Haar-like, các hàm cơ bản trong OpenCV ứng dụng trong phát hiện khuôn mặt trong ảnh.
- Về mặt thực nghiệm: Đồ án đã cài đặt được chương trình phát hiện khuôn mặt trong ảnh sử dụng các hàm của OpenCV, và mới dừng lại ở phát hiện khuôn mặt nhìn thẳng (độ nghiêng, xoay ít).
- Về hình thức: Báo cáo trình bày sáng sủa, bố cục hợp lý.
- Đồ án đáp ứng được yêu cầu đề ra.

3. Ý kiến của cán bộ hướng dẫn:

Đạt Không đạt Điểm:.....

Ngày 25 tháng 9 năm 2019
Cán bộ hướng dẫn

TS. Ngô Trường Giang

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập - Tự do - Hạnh phúc

PHIẾU NHẬN XÉT CỦA CÁN BỘ CHẤM PHẢN BIỆN

Họ và tên giảng viên: TS. Đỗ Văn Chiêu.

Đơn vị công tác: Khoa Công nghệ Thông tin – Trường Đại học Dân lập Hải Phòng

Họ và tên sinh viên: Cao Tiến Đạt **Ngành:** Công nghệ Thông tin

Đề tài tốt nghiệp:

“Xây dựng ứng dụng phát hiện khuôn mặt trong ảnh sử dụng OpenCV”

1. Phân nhận xét của giảng viên chấm phản biện

.....
.....
.....
.....
.....
.....
.....
.....
.....
.....

2. Những mặt còn hạn chế

.....
.....
.....
.....
.....
.....

3. Ý kiến của giảng viên chấm phản biện

Được bảo vệ **Không được bảo vệ** **Điểm:**.....

Hải Phòng, ngày tháng 10 năm 2019

Cán bộ chấm phản biện

(Ký và ghi rõ họ tên)

LỜI CẢM ƠN

Em xin chân thành cảm ơn các thầy, các cô khoa Công nghệ Thông tin Trường Đại học quản lí và công nghệ đã tình dạy dỗ, truyền đạt cho chúng em nhiều kiến thức quý báu.

Em xin tỏ lòng biết ơn sâu sắc đến thầy TS.Ngô Trường Giang, người đã tận tình giúp đỡ và truyền đạt nhiều kinh nghiệm để đề tài có thể được thực hiện và hoàn thành.

Xin chân thành cảm ơn các bạn trong khoa Công Nghệ Thông Tin, Đại Học Quản lí và công nghệ đã giúp đỡ mình rất nhiều trong quá trình thực hiện đề tài.

Em xin trân trọng cảm ơn!

Hải Phòng, tháng 9 năm 2019.

Sinh viên

Cao Tiến Đạt

MỤC LỤC

PHẦN MỞ ĐẦU	5
CHƯƠNG 1: Tổng quan về phát hiện khuôn mặt.	6
1.1 Giới thiệu phát hiện khuôn mặt	6
1.2 Các hướng tiếp cận trong phát hiện khuôn mặt	7
1.2.1 Hướng tiếp cận dựa trên tri thức	8
1.2.2 Hướng tiếp cận dựa trên đặc trưng không đổi.	11
1.2.3 Hướng tiếp cận dựa trên so lớp mẫu	15
1.2.4 Hướng tiếp cận dựa trên máy học	18
1.2.5 Hướng tiếp cận tổng hợp.....	23
1.3 Khó khăn và thách thức trong bài toán phát hiện khuôn mặt	24
1.4 Một số lĩnh vực ứng dụng phát hiện khuôn mặt	25
CHƯƠNG 2: Phát hiện khuôn mặt sử dụng đặc trưng Haar	26
2.1 Đặc trưng của Haar	26
2.1.1 Đối tượng nhận dạng.....	26
2.1.2 Đặc trưng.....	27
2.2 Bộ phân lớp AdaBoost.....	30
2.2.1 Giới thiệu.....	30
2.2.2 Các hướng tiếp cận dò tìm khuôn mặt nhanh	30
2.2.3 Thuật toán AdaBoost	31
2.2.4 Bộ phân lớp AdaBoost.....	36
2.3 Dò tìm khuôn mặt	39
2.4 Nhận xét	40
2.4.1 Ưu điểm.....	40
2.4.2 Nhược điểm.....	41
CHƯƠNG 3: Thực nghiệm phát hiện khuôn mặt	42
3.1 Mô tả bài toán	42
Phân tích và lựa chọn công cụ.....	42
3.2 Phát hiện khuôn mặt sử dụng hàm OpenCV.....	43
3.2.1 OpenCV là gì.....	43
3.2.2 Cấu trúc OpenCV	44

3.2.3 Phát hiện khuôn mặt với các hàm trong OpenCV	45
3.3 Một số kết quả chương trình	49
KẾT LUẬN	51

DANH MỤC HÌNH

Hình 1-1: Các bước chính trong một hệ thống nhận dạng mặt người	6
Hình 2-1.....	27
Hình 2-2: 4 đặt trung Haar-like cơ bản.....	27
Hình 2-3: Đặc trưng đường(line feature)	28
Hình 2-4: Đặc trưng xung quanh tâm(center-surround features)	28
Hình 2-5: Các đặc trưng mở rộng của các đặc trưng Haar-like cơ sở	28
Hình 2-6: Công thức tính Intergral Image	29
Hình 2-7: Ví dụ cách tính nhanh các giá trị mức xám của vùng D trên ảnh ..	30
Hình 2-8: Ví dụ minh họa cho thuật toán AdaBoost	36
Hình 2-9: Minh họa bộ đồ tìm phân tầng.....	37
Hình 3-1: Hệ thống phát hiện khuôn mặt người	42
Hình 3-2: Cấu trúc cơ bản của OpenCV	44
Hình 3-3: Trong thư mục của Opencv đã có các bộ huấn luyện có sẵn.	45
Hình 3-4: Chuyên về ảnh xám	46
Hình 3-5: Tầm nhìn của máy tính khi các đặc trưng Haar chạy	47
Hình 3-6.....	47
Hình 3-7: Các đặc trưng Haar đánh dấu những nơi có thể là khuôn mặt	48
Hình 3-8: Bức ảnh đã được xác định khuôn mặt	49
Hình 3-9: Bức ảnh cần nhận diện khuôn mặt.....	49
Hình 3-10: Giao diện chương trình	50
Hình 3-11: Nhập đường dẫn tới ảnh cần nhận diện	50
Hình 3-12: Kết quả.....	50
Hình 3-13: Bức ảnh có nhiều khuôn mặt	51
Hình 3-14: Kết quả (khuôn mặt nghiêng chưa nhận dạng được).....	51

PHẦN MỞ ĐẦU

Công nghệ thông tin đang được ứng dụng trong mọi lĩnh vực của cuộc sống. Với một hệ thống máy tính, chúng ta có thể làm được rất nhiều việc, tiết kiệm thời gian và công sức. Điển hình như công việc nhận dạng mặt người. Ngày xưa, muốn tìm kiếm một kẻ tình nghi trong siêu thị hay sân bay, các nhân viên an ninh phải tìm kiếm trên từng màn hình camera theo dõi. Ngày nay, công việc đấy đã được làm tự động nhờ các hệ thống nhận dạng mặt người. Phát hiện mặt người trong ảnh là một phần quan trọng của hệ thống nhận dạng mặt người đó, giải quyết tốt việc phát hiện mặt người sẽ giúp tiết kiệm thời gian và nâng cao độ chính xác của việc nhận dạng khuôn mặt.

Phát hiện mặt người cũng là một bài toán nhận dạng đơn giản, hệ thống chỉ cần phân loại đối tượng đưa vào có phải mặt người hay không phải mặt người. Ở mức độ cao hơn, sau khi đã phát hiện được khuôn mặt, các khuôn mặt đó sẽ được so sánh với các khuôn mặt có trong dữ liệu để nhận dạng xem khuôn mặt đấy là của ai (thường áp dụng trong nhận dạng khuôn mặt của người nổi tiếng hoặc của tội phạm đang bị truy nã).

Với mục tiêu chính là tìm hiểu đặc trưng haar-like, mô hình Cascade of Classifiers, đồng thời áp dụng vào bài toán phát hiện mặt người trong ảnh.

Nội dung đồ án bao gồm:

Chương 1: Tổng quan về phát hiện khuôn mặt.

Chương 2: Phát hiện khuôn mặt sử dụng đặc trưng Haar.

Chương 3: Thực nghiệm phát hiện khuôn mặt.

Phần kết luận.

Tài liệu tham khảo.

CHƯƠNG 1: TỔNG QUAN VỀ PHÁT HIỆN KHUÔN MẶT.

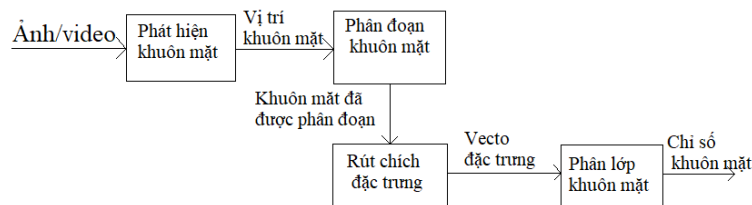
1.1 Giới thiệu phát hiện khuôn mặt

Trong những năm gần đây, có rất nhiều công trình nghiên cứu về bài toán xác định khuôn mặt người từ ảnh đen trắng, xám đến ảnh màu. Ban đầu chỉ là những bài toán đơn giản, mỗi ảnh chỉ có một khuôn mặt nhìn thẳng và đầu luôn phải ở tư thế thẳng đứng trong ảnh đen trắng, không đáp ứng được nhu cầu ngày càng cao trong cuộc sống, khoa học ngày nay. Vì thế đã có những cải tiến nghiên cứu về bài toán phát hiện khuôn mặt người trong những môi trường phức tạp hơn, có nhiều khuôn mặt người trong ảnh hơn, và có nhiều tư thế thay đổi trong ảnh.

Xác định khuôn mặt người là một kỹ thuật để xác định vị trí và kích thước khuôn mặt người trong các ảnh bất kỳ. Kỹ thuật này nhận biết về các đặc trưng của khuôn mặt và bỏ qua những thứ khác.

Tổng quan kiến trúc của một hệ thống nhận dạng mặt người

Một hệ thống nhận dạng mặt người thông thường bao gồm bốn bước xử lý sau: phát hiện khuôn mặt (face detection), phân đoạn khuôn mặt (face alignment hay segmentation), rút trích đặc trưng (feature extraction), và phân lớp khuôn mặt (face classification).



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

Phát hiện khuôn mặt dò tìm và định vị những vị trí khuôn mặt xuất hiện trong ảnh hoặc trên các frame video. Phân đoạn khuôn mặt sẽ xác định vị trí mắt mũi, miệng, là các thành phần khác của khuôn mặt và chuyển kết quả này cho bước rút trích đặc trưng. Từ những thông tin về các thành phần trên khuôn mặt, chúng ta có thể dễ dàng tính được véc-tơ đặc trưng trong bước rút trích đặc trưng. Những véc-tơ đặc trưng này sẽ là dữ liệu đầu vào cho một mô hình đã được huấn luyện trước để phân loại khuôn mặt. Bên cạnh những bước chính nêu trên, chúng ta còn có thể áp dụng thêm một số bước khác như tiền xử lý, hậu xử lý nhằm làm tăng độ chính xác cho hệ thống. Do một số thông số như: tư thế khuôn mặt, độ sáng, điều kiện ánh sáng, v.v , phát hiện khuôn mặt được đánh giá là bước khó khăn và quan trọng nhất so với các bước còn lại của hệ thống. Trong luận văn này, em tập trung chủ yếu vào bước phát hiện khuôn mặt.

1.2 Các hướng tiếp cận trong phát hiện khuôn mặt

Có nhiều nghiên cứu tìm ra phương pháp xác định khuôn mặt người, từ ảnh xám đến ngày nay là ảnh màu. Dựa vào tính chất của 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 các loại khuôn mặt người thành các luật. Thông thường các luật mô tả quan hệ của các đặc trưng.
- **Hướng tiếp cận dựa trên đặc trưng không thay đổi:** Mục tiêu 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, vị trí đặt thiếu bị thu hình hoặc điều kiện ánh sáng thay đổi.
- **Hướng tiếp cận dựa trên so khớp mẫu:** Dùng các mẫu chuẩn của khuôn mặt người(các mẫu này được chọn lựa và lưu trữ) để mô tả cho khuôn mặt người hay các đặc trưng khuôn mặt(các mẫu

này phải chọn làm sao cho tách biệt nhau theo tiêu chuẩn mà các tác giả định ra để so sánh).

- **Hướng tiếp cận dựa trên diện mạo:** Trái ngược hẳn so khớp mẫu, các mô hình học ở đây được học từ 1 tập ảnh huấn luyện cho trước. Sau đó hệ thống sẽ xác định khuôn mặt người. Một số tác giả còn gọi hướng tiếp cận này là hướng tiếp cận theo phương pháp học.

1.2.1 Hướng tiếp cận dựa trên tri thức

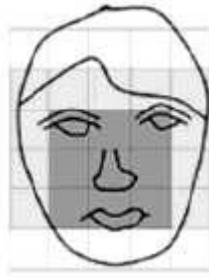
Trong hướng tiếp cận này, các luật sẽ phụ thuộc rất hơn vào tri thức của những tác giả nghiên cứu về bài toán xác định khuôn mặt người. Đây là hướng tiếp cận top-down. Dễ dàng xây dựng các luật cơ bản để mô tả các đặc trưng của khuôn mặt và các quan hệ tương ứng, Ví dụ, một khuôn mặt thường có hai đôi 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 các tác giả sẽ trích đặc trưng của khuôn mặt trước tiên để có thể 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 là 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 những khuôn mặt(chặt chẽ) 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 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 khi cần mở rộng yêu cầu của bài toán để xác định các khuôn mặt có nhiều tư thế khác nhau.



Hình 1-2: (a) Ảnh ban đầu có độ phân giải $n = 1$; (b), (c), và (d) Ảnh có độ phân giải $n = 4, 8, 16$.

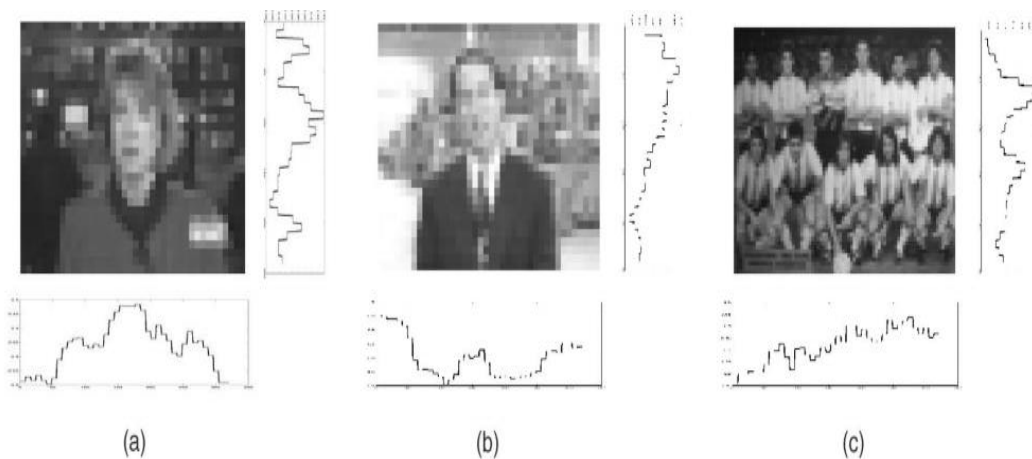
Yang và Huang dùng phương thức theo hướng tiếp cận này để xác định các khuôn mặt. Hệ thống của hai tác giả này bao gồm ba mức luật. Ở mức cao nhất, dùng một khung cửa sổ quét trên ảnh và thông qua một tập luật để tìm các ứng viên có thể là khuôn mặt. Ở mức kế tiếp, hai ông dùng một tập luật để mô tả tổng quát hình dáng khuôn mặt. Ở mức còn 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, (Hình 1.1). Các luật ở mức cao nhất để tìm ứng viên như: “Vùng trung tâm khuôn mặt (phần tối hơn trong hình 1.2) có bốn phần với một mức độ đều cơ bản”, “phần xung quanh bên trên của một khuôn mặt (phần sáng hơn trong hình 1.2) 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”. Ở mức hai, xem xét biểu đồ của các ứng viên để loại bớt ứng viên 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.



Hình 1-3: Một loại tri thức của người nghiên cứu phân tích trên khuôn mặt.

Kotropoulos và Pitas đưa ra một phương pháp 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 $m \times n$ iwr tại vị trí (x,y) , các hàm để chiếu ảnh theo phương ngang và thẳng đứng được định nghĩa như sau:

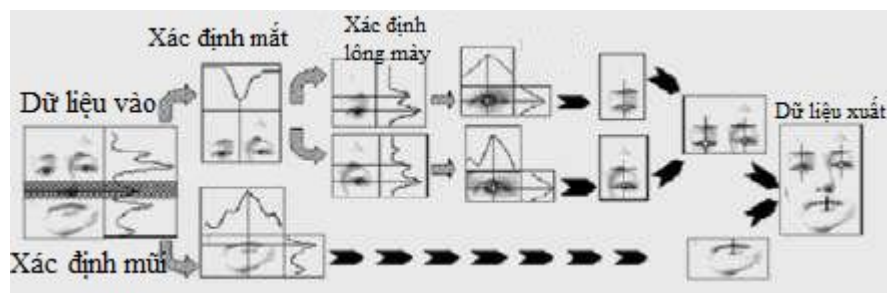
$$HI(x)=\sum_{ny-1} I(x, y) \text{ và } VI(y)=\sum_{mx-1} I(x, y) \quad (1-1)$$



Hình 1-4: 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

Dựa trên biên đồ chiếu hình ngang, có hai cực tiểu cục bộ khi hai ông xét quá trình thay đổi độ dốc HI, đó chính là cạnh bên trái và phải của hai bên đầu. Tương tự với hình chiếu dọc VI, các cực tiểu cục bộ cũng cho ta biết vị trí miệng, đỉnh mũi, và hai mắt. Các đặc trưng này đủ để xác định khuôn mặt. Hình 1.4.a là một ví dụ của cách xác định trên. Cách xác định này có tỷ lệ xác định chính xác 86,5% cho trường hợp chỉ có một khuôn mặt thẳng trong ảnh và hình nền không phức tạp. Nếu hình nền phức tạp thì rất khó tìm, như là hình 1.4.b. Nếu ảnh có nhiều khuôn mặt thì sẽ không xác định được, hình 1.4.c.



Hình 1-5: Chiếu từng phần ứng viên để xác định khuôn mặt.

Mateos và Chicote dùng kết cấu để xác định ứng viên trong ảnh màu. Sau đó phân tích hình dáng, kích thước, thành phần khuôn mặt để xác định khuôn mặt, hai ông trích các ứng viên của từng thành phần khuôn mặt, sau đó chiếu từng phần này để xác thực đó có phải là thành phần khuôn mặt hay không. Tỷ lệ chính xác hơn 87%.

Berbar kết hợp mô hình màu da người và xác định cạnh để tìm ứng viên khuôn mặt người. Sau đó kết hợp các đặc trưng và phương pháp chiếu các ứng viên khuôn mặt xuống hệ trục tọa độ để xác định ứng viên nào thực sự là khuôn mặt.

1.2.2 Hướng tiếp cận dựa trên đặc trưng không đổi.

Đây là hướng tiếp cận thiếu 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 trong

các tư thế khác nhau và điều kiện ánh sáng khác nhau. Do đó khuôn mặt phải có các thuộc tính hay đặc trưng không thay đổi. Theo nhiều nghiên cứu thì ban đầu phải xác định các đặc trưng khuôn mặt rồi chỉ 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 thông 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 theo cạnh mới, mà cạnh này lại rõ hơn cạnh thực sự của khuôn mặt, vì thế nếu dùng cạnh để xác định sẽ gặp khó khăn.

1.2.2.1 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. Graf đưa ra một phương pháp xác định đặc trưng rồi xác định khuôn mặt trong ảnh xám. Dùng bộ lọc để làm nổi các biên, các phép toán hình thái học (morphology) được dùng để làm nổi bật các vùng có cường độ cao và hình dáng chắc chắn (như mắt). Thông qua histogram để tìm các đỉnh nổi bật để xác định các ngưỡng chuyển ảnh xám

thành hai ảnh nhị phân. Các thành phần dính nhau đều xuất hiện trong hai ảnh nhị phân thì được xem là vùng của ứng viên khuôn mặt rồi phân loại xem có phải là khuôn mặt không. Phương pháp được kiểm tra trên các ảnh chỉ có đầu và vai của người. Tuy nhiên còn có một vấn đề ở đây là làm sao để sử dụng các phép toán morphology và làm sao xác định khuôn mặt trên các vùng ứng viên.

1.2.2.2 Đặc trưng kết cấu

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). Có ba loại đặc trưng được xem xét: màu da, tóc, và những thứ khác. Hai ông dùng mạng nơ-ron về môi 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. Dai và Nakano dùng mô hình SGLD để xác định khuôn mặt người. Thông tin màu sắc được kết hợp với mô hình kết cấu khuôn mặt. Hai tác giả xây dựng thuật giải xác định khuôn mặt trong không gian màu, với các phần tựa màu cam để xác định các vùng có thể là khuôn mặt người.

Một thuận lợi của phương pháp này là có thể xác định khuôn mặt không chỉ chụp thẳng và có thể có râu và có kính. Mark và Andrew dùng phân bố màu da và thuật toán DoG (Difference of Gauss) để tìm các ứng viên, rồi xác thực bằng một hệ thống học kết cấu của khuôn mặt. 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 thông 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%.

1.2.2.3 Đặc trưng 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.

1.2.2.4 Đa đặc trưng

Gần đây có nhiều nghiên cứu sử dụng các đặc trưng toàn cục (tổng quát) 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 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. Yachida đưa ra một phương pháp xác định khuôn mặt người trong ảnh màu bằng lý thuyết logic mờ.

Ông dùng hai mô hình mờ để mô tả phân bố màu da người và màu tóc trong không gian màu CIE XYZ. Năm mô hình hình dạng của đầu (một thẳng và bốn xoay xung quanh) để mô tả hình dáng của mặt trong ảnh. Mỗi mô hình hình dạng là một mẫu 2-chiều bao gồm các ô vuông có kích thước $m \times n$, mỗi ô có thể chứa nhiều hơn một điểm ảnh. Hai thuộc tính được gán cho mỗi ô là: tỷ lệ màu da và tỷ lệ tóc, chỉ ra tỷ lệ diện tích vùng da (tóc) trong ô so với diện tích của ô. Mỗi điểm ảnh sẽ được phân loại thành tóc, khuôn mặt, tóc/khuôn mặt, và tóc/nền trên cơ sở phân bố của mô hình, theo cách đó sẽ có được các vùng giống khuôn mặt và giống tóc. Mô hình hình dáng của đầu sẽ được so sánh với vùng giống khuôn mặt và giống tóc. Nếu tương tự, vùng đang xét sẽ trở thành ứng viên khuôn mặt, sau đó dùng các đặc trưng mátlông mày và mũi-miệng để xác định ứng viên nào sẽ là khuôn mặt thật sự.

Sobottka và Pitas dùng các đặc trưng về hình dáng và màu sắc để xác định khuôn mặt người. Dùng một ngưỡng để phân đoạn trong không gian màu HSV để xác định các vùng có thể là màu da người (vùng giống màu da người). Các thành phần dính nhau sẽ được xác định bằng thuật toán tăng vùng

ở độ phân giải thô. Xem xét tiền ứng viên nào vừa khớp hình dạng ellipse sẽ được chọn làm ứng viên của khuôn mặt. Sau đó dùng các đặc trưng bên trong như: mắt và miệng, được trích ra trên cơ sở các vùng mắt và miệng sẽ tốt hơn các vùng khác của khuôn mặt, sau cùng phân loại dựa trên mạng nơ-ron để biết vùng ứng viên nào là khuôn mặt người và vùng nào không phải khuôn mặt người. Tỷ lệ chính xác là 85%.

1.2.3 Hướng tiếp cận dựa trên so lớp mẫu

Trong so khớp mẫu, các 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, 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 rất dễ cài đặt, nhưng không hiệu quả khi có sự thay đổi về tỷ lệ, tư thế và hình giáng.

1.2.3.1 Xác định mẫu trước

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.

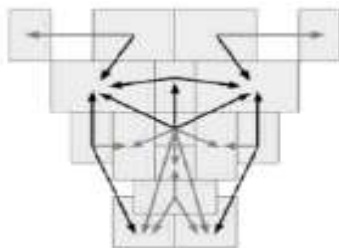
Craw đưa ra một phương pháp xác định khuôn mặt người dựa vào các mẫu về hình dáng của các ảnh được chụp thẳng (dùng về bề ngoài của hình dáng khuôn mặt). Đầu tiên dùng phép lọc Sobel để tìm các cạnh. Các cạnh

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.

Sinha dùng một tập nhỏ các bất biến ảnh trong không gian ảnh để mô tả không gian các mẫu ảnh. 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. 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. Các vùng có độ sáng đều được xem như một mẫu tỷ số mà là mẫu thô trong không gian ảnh của một khuôn mặt với độ thích hợp ít dùng để chọn như các đặc trưng chính của khuôn mặt như hai mắt, hai má, và trán. Lưu giữ 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 khi một ảnh thỏa tất cả các cặp sáng hơn – tối hơn. Ý tưởng này xuất phát từ sự khác biệt của cường độ giữa các vùng kề cục bộ, sau này được mở rộng trên cơ sở biến đổi wavelet để biểu diễn cho xác định người đi bộ, xác định xe hơi, xác định khuôn mặt. Ý tưởng của Sinha còn được áp dụng cho hệ thống thị giác của robot. Hình 1-5 cho thấy mẫu nổi bật trong 23 quan hệ được định nghĩa. Dùng các quan hệ này để phân loại, có 11 quan hệ thiết yếu (các mũi tên màu đen) và 12 quan hệ xác thực (các mũi tên xám). Mỗi mũi tên là một quan hệ. Một quan hệ thỏa mãn mẫu khuôn mặt khi tỷ lệ giữa hai vùng vượt qua một ngưỡng và 23 quan hệ này vượt ngưỡng thì xem như xác định được một khuôn mặt.

Phương pháp so khớp mẫu theo thứ tự để xác định khuôn mặt người do Miao trình bày. Ở giai đoạn đầu tiên, ảnh sẽ được xoay từ -20o đến 20o với

mỗi bước là 50 và theo thứ tự. Xây dựng ảnh đa độ phân giải, hình 1-1, rồi dùng phép toán Laplace để xác định các cạnh. Một mẫu khuôn mặt gồm các cạnh mô tả sáu thành phần: hai lông mày, hai mắt, một mũi, và một miệng. Sau đó áp dụng heuristic để xác định sự tồn tại của khuôn mặt trong ảnh, phương pháp này cho phép xác nhiều khuôn mặt, nhưng kết quả không tốt bằng xác định một khuôn mặt (chụp thẳng hoặc xoay) trong ảnh xám.



Hình 1-6: Mẫu khuôn mặt, có 16 vùng và 23 quan hệ (các mũi tên).

Wei và Lai dùng bộ lọc để phân đoạn kết hợp thuật toán tìm láng giềng gần nhất xác định ứng viên khuôn mặt, từ ứng viên này sau đó so khớp với các mẫu đã xác định trước để biết ứng viên có phải là khuôn mặt hay không. Tỷ lệ chính xác là 80%.

Darrell dùng phân đoạn để tìm ứng viên, dùng ứng viên này để xác định khuôn mặt người dựa vào mẫu rồi theo vết chuyển động của người.

Dowdall dùng phổ của màu da người để xác định ứng viên. Sau đó chiếu các ứng viên này để so sánh với các mẫu có trước để xác định ứng viên nào là khuôn mặt người. Phương pháp này chỉ xác định cho khuôn mặt chụp thẳng và gần thẳng, góc quay khoảng từ -10o đến 10o.

1.2.3.2 Các mẫu bị biến dạng

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.

Lanitis mô tả một phương pháp biểu diễn khuôn mặt người với cả hai thông tin: hình dáng và cường độ. Bắt đầu với các tập ảnh được huấn luyện với các đường viền mẫu như là đường bao mắt, mũi, cằm/má được gán nhãn. Dùng một vector các điểm mẫu để mô tả hình dáng. Tác giả dùng một mô hình phân bố điểm (Point Distribution Model – PDM) để mô tả vector hình dáng qua toàn bộ các cá thể.

1.2.4 Hướng tiếp cận dựa trên máy học

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 giả 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 $p(x | \text{khuôn mặt})$ và $p(x | \text{không phải khuôn mặt})$. Có thể dùng phân loại Bayes.

1.2.4.1 Eigenface

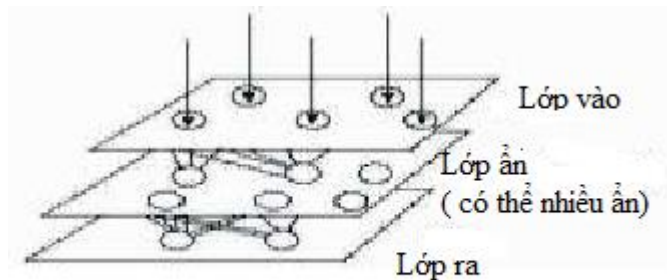
Kohonen đã đưa ra phương pháp dùng vector riêng để nhận dạng khuôn mặt, ông dùng một mạng nơ-ron đơn giản để chứng tỏ khả năng của phương pháp này trên các ảnh đã được chuẩn hóa. Mạng nơ-ron tính một mô tả của khuôn mặt bằng cách xấp xỉ các vector riêng của ma trận tương quan của ảnh. Các vector riêng sau này được biết đến với cái tên Eigenface. Kirby và Sirovich chứng tỏ các ảnh có các khuôn mặt có thể được mã hóa tuyến tính bằng một số lượng vừa phải các ảnh cơ sở. Tính chất này dựa trên biến đổi KarhunenLòeve, mà còn được gọi dưới một cái tên khác là PCA và biến đổi Hotelling. Ý tưởng này được xem là của Pearson trình bày đầu tiên vào năm 1901 và sau đó là Hotelling vào năm 1933. Cho một tập các ảnh huấn luyện có kích thước $n \times m$ được mô tả bởi các vector có kích thước $m \times m$, các vector cơ sở cho một không gian con tối ưu được xác định thông qua lỗi bình phương trung bình khi chiếu các ảnh huấn luyện vào không gian con này. Các tác giả gọi tập các vector cơ sở tối ưu này là ảnh riêng sau đó gọi cho đơn giản là vector riêng của ma trận hiệp phương sai được tính từ các ảnh khuôn mặt đã vector hóa trong tập huấn luyện. Nếu cho 100 ảnh, mà mỗi khuôn mặt có kích thước 91×50 thì có thể chỉ dùng 50 ảnh riêng, trong khi vẫn duy trì được một khả năng giống nhau hợp lý (giữ được 95% tính chất).

Turk và Pentland áp dụng PCA để xác định và nhận dạng khuôn mặt. Tương tự, dùng PCA trên tập huấn luyện ảnh các khuôn mặt để sinh các ảnh riêng (còn gọi là eigenface) để tìm một không gian con (không gian khuôn mặt) trong không gian ảnh. Các ảnh khuôn mặt được chiếu vào không gian con này và được gom nhóm lại. Tương tự các ảnh không có khuôn mặt dùng để huấn luyện cũng được chiếu vào cùng không gian con và gom nhóm lại. Các ảnh khi chiếu vào không gian khuôn mặt thì không bị thay đổi tính chất cơ bản, trong khi chiếu các ảnh không có khuôn mặt thì xuất hiện sự khác nhau cũng không ít. Xác định sự có mặt của một khuôn mặt trong ảnh thông qua tất cả khoảng cách giữa các vị trí trong ảnh và không gian ảnh. Khoảng

cách này dùng để xem xét có hay không có khuôn mặt người, kết quả khi tính toán các khoảng cách sẽ cho ta một bản đồ về khuôn mặt. Có thể xác định được từ cực tiểu địa phương của bản đồ này. Có nhiều nghiên cứu về xác định khuôn mặt, nhận dạng, và trích đặc trưng từ ý tưởng vector riêng, phân rã, và gom nhóm. Sau đó Kim phát triển cho ảnh màu, bằng cách phân đoạn ảnh để tìm ứng đề không gian tìm kiếm bớt đi.

1.2.4.2 Phương pháp Mạng neuron

Mô phỏng hoạt động của các nơ-ron thần kinh, mạng nơ-ron nhân tạo là hệ thống bao gồm nhiều phần tử xử lý đơn giản (neuron) hoạt động song song. Tính năng của hệ thống này tùy thuộc vào cấu trúc của hệ, các trọng số liên kết nơ-ron và quá trình tính toán tại các nơ-ron đơn lẻ. Mạng nơ-ron có thể từ dữ liệu mẫu và tổng quát hoá dựa trên các dữ liệu mẫu học.



Hình 1-7: Mô hình mạng Neuron

Một nhóm các nơ-ron được tổ chức theo một cách sao cho tất cả chúng đều nhận cùng một vector vào X để xử lý tại cùng một thời điểm. Việc sản sinh ra tín hiệu ra của mạng xuất hiện cùng một lúc. Vì mỗi nơ-ron có một tập trọng số khác nhau nên có bao nhiêu nơ-ron sẽ sản sinh ra bấy nhiêu tín hiệu ra khác nhau. Một nhóm các nơ-ron như vậy được gọi là một lớp mạng. Chúng ta có thể kết hợp nhiều lớp mạng tạo ra một mạng có nhiều lớp, lớp nhận tín hiệu đầu vào (vector tín hiệu vào x) được gọi là lớp vào (input layer). Trên thực tế chúng thực hiện như một bộ đệm chứa tín hiệu đầu vào. Các tín

hiệu đầu ra của mạng được sản sinh ra từ lớp ra của mạng (output layer). Bất kỳ lớp nào nằm giữa 2 lớp mạng trên được gọi là lớp ẩn (hidden layer) và nó là thành phần nội tại của mạng và không có tiếp xúc nào với môi trường bên ngoài. Số lượng lớp ẩn có thể từ 0 đến vài lớp. Mô hình nơ-ron nhân tạo đòi hỏi 3 thành phần cơ bản sau:

- Tập trọng số liên kết đặc trưng cho các khớp thần kinh.
- Bộ cộng (Sum) để thực hiện phép tính tổng các tích tin hiệu vào với trọng số liên kết tương ứng.
- Hàm kích hoạt (squashing function) hay hàm chuyển (transfer function) thực hiện giới hạn đầu vào của neuron.

Trong mô hình nơ-ron nhân tạo mỗi nơ-ron được nối với các nơ-ron khác và nhận được tín hiệu xi từ chúng với các trọng số wi. Tổng thông tin vào có trọng số là: $Net = \sum w_i x_i$.

1.2.4.3 Phương pháp SVM – support vector machine

SVM là phương pháp do Vladimir N. Vapnik đề xuất năm 1995. SVM dựa trên lý thuyết thống kê và ngày càng được sử dụng phổ biến trong nhiều lĩnh vực, đặc biệt là trong lĩnh vực phân loại mẫu và nhận dạng mẫu. Đồng thời phương pháp này có nhiều tính năng ưu việt so với các phương pháp cổ điển khác như dễ dàng xử lý, xử lý có tính ổn định cao trên dữ liệu phức tạp, có thể có số chiều lớn và quan trọng hơn cả là khả năng xử lý tổng quát.

1.2.4.4 Mô hình Markov ẩn

Phương pháp tìm kiếm khuôn mặt dựa trên mô hình Markov ẩn cũng là một trong các hướng nghiên cứu được chú trọng. Mô hình Markov ẩn là một tập các mô hình thống kê được sử dụng để mô tả các đặc tính thống kê của tín hiệu. Lý thuyết về chuỗi Markov và mô hình Markov đã được nghiên cứu sâu rộng và áp dụng nhiều trong lý thuyết nhận dạng như nhận dạng tiếng nói, chữ viết. Samarie và cộng sự, Netfian và cộng sự là hai nhóm nghiên cứu

hàng đầu trong việc áp dụng mô hình Markov ẩn vào tìm kiếm và nhận dạng mặt người.

1.2.4.5 Mạng lọc thưa (Sparse Network of Winnows – SNoW)

Đây là phương pháp do Yang đề xuất dùng để xác định khuôn mặt người với các đặc trưng khác nhau và biểu diễn trong các tư thế khác nhau, dưới điều kiện ánh sáng khác nhau. SNoW là một mạng thưa dùng các hàm tuyến tính và dùng lọc để cập nhật luật. Phương pháp này thích hợp cho học trong miền khi các đặc trưng tiềm năng tạo ra các quyết định sai khác nhau mà không biết mức độ ưu tiên. Tỷ lệ lỗi là 5.9%, hiệu quả cũng như các phương pháp khác.

1.2.4.6 AdaBoost

AdaBoost là một phân loại mạnh phi tuyến phức $HM(x)$, được xây dựng từ M phân loại yếu. Mục tiêu của AdaBoost là học một dãy các phân loại yếu. Giả sử có một tập N mẫu huấn luyện đã được gán nhãn $\{(x_1, y_1), \dots, (x_N, y_N)\}$, với y_i là nhãn tương ứng của mẫu $x_i = \mathbf{R}^n$. Tính một phân bố của các mẫu huấn luyện $[w_1, \dots, w_N]$ cập nhật trong suốt quá trình học. Sau bước m , mẫu khó phân loại (x_i, y_i) có trọng số mới $w_i^{(m)}$, đến bước lặp thứ $(m+1)$, mẫu này sẽ có tầm quan trọng hơn. 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.

1.2.4.7 Học với FloatBoost

Li và Zhang đưa ra một khái niệm mới đó là FloatBoost. Phương pháp này học dựa trên phân loại boosting để tỷ lệ lỗi cực tiểu. Nhưng phương pháp này cho phép quay lui sau khi tại mỗi bước chẵn khi học bằng Adaboost đã cực tiểu được tỷ lệ lỗi trực tiếp, cực tiểu theo hàm mũ. Có hai vấn đề gặp khi dùng phương pháp Adaboost:

- Thứ nhất: AdaBoost cực tiểu theo hàm mũ tại biên qua tập huấn luyện. Đây là tiện lợi, tuy nhiên mục tiêu cuối cùng trong các ứng dụng dùng phân loại mẫu thì thường là cực tiểu một giá trị trực

tiếp (tuyến tính) kết hợp với tỷ lệ lỗi. Một phân loại cực mạnh được học bằng AdaBoost thì gần điểm tối ưu của ứng dụng trong điều kiện tỉ lệ lỗi.

- Thứ hai: AdaBoost để lại một thách thức nếu dùng phân loại yếu để học, Học để phân loại tối ưu khi dùng phân loại yếu cần ước lượng mật độ không gian đặc trưng, điều này là vấn đề khó, đặc biệt khi số chiều của không gian khá lớn.

Một thuật toán yếu có hiệu quả và dễ dùng thì rất cần thiết. Float xem như một cầu nối giữa mục tiêu của học boosting thông thường (cực đại biên) và ứng dụng dùng cực tiểu tỷ lệ lỗi thông qua việc kết hợp phương pháp tìm kiếm Floating và AdaBoost kết hợp kỹ thuật quay lui.

1.2.5 Hướng tiếp cận tổng hợp

Các phương pháp được chia làm bốn phân loại chính theo bốn hướng tiếp cận. Tuy nhiên, có nhiều phương pháp không hoàn toàn rơi vào một trong bốn hướng tiếp cận này mà ở trong bốn hướng tiếp cận khác nhau. Ví dụ, phương pháp so khớp mẫu dùng mô hình khuôn mặt người và các mẫu con để trích các đặc trưng khuôn mặt, và sau đó dùng các đặc trưng này để xác định khuôn mặt. Hơn nữa phương pháp dựa trên tri thức và phương pháp so khớp mẫu không thực sự tách biệt, từ đó có nhiều hướng giải quyết dùng tri thức của con người để định nghĩa các mẫu khuôn mặt người.

Kim kết hợp các đặc trưng láng giềng của khuôn mặt xây dựng các mẫu theo các hướng, sau đó dùng kỹ thuật xác định cạnh EBM (Edge-like Blob Map) theo cường độ. Ông xây dựng logic mờ kết hợp PCA để ước lượng tư thế các khuôn mặt.

Taur và Tao xây dựng phân loại neurofuzzy (neuro-fuzzy classifier – NEFCAR) có độ do tin cậy để biết ảnh nào là khuôn mặt người. Các ứng viên được chọn thông qua phân đoạn màu da.

1.3 Khó khăn và thách thức trong bài toán phát hiện khuôn mặt

Việc xác định khuôn mặt người có những khó khăn nhất định:

- Hướng (pose) của khuôn mặt đối với máy ảnh, như: nhìn thẳng, nhìn nghiêng hay nhìn từ trên xuống. Cùng trong một ảnh có thể có nhiều khuôn mặt ở những tư thế khác nhau.
- Sự có mặt của các chi tiết không phải là đặc trưng riêng của khuôn mặt người, như: râu quai nón, mắt kính,
- Các nét mặt (facial expression) khác nhau trên khuôn mặt, như: vui, buồn, ngạc nhiên,
- Mặt người bị che khuất bởi các đối tượng khác có trong ảnh.
- Điều kiện ảnh, đặc biệt là về độ sáng và chất lượng ảnh, chất lượng thiết bị thu hình.
- Trục tọa độ của máy ảnh so với ảnh.
- Kích thước khác nhau của các khuôn mặt người, và đặc biệt là trong cùng một ảnh.
- Nhiều khuôn mặt có vùng da dính lẫn nhau.

Các khó khăn trên chứng tỏ rằng bất cứ phương pháp giải quyết (thuật toán) bài toán xác định khuôn mặt người nào cũng sẽ không thể tránh khỏi một số khiếm khuyết nhất định. Để đánh giá và so sánh các phương pháp xác định mặt người, người ta thường dựa trên các tiêu chí sau:

- Tỷ lệ xác định chính xác là tỷ lệ số lượng các khuôn mặt người được xác định đúng từ hệ thống khi sử dụng một phương pháp để xây dựng so với số lượng khuôn mặt người thật sự có trong các ảnh (detection rate).

- Số lượng xác định nhầm là số lượng vùng trong ảnh không phải là khuôn mặt người mà hệ thống xác định nhầm là khuôn mặt người (false positives).
- Thời gian thực hiện là thời gian để máy tính xác định khuôn mặt người trong ảnh (running time).

1.4 Một số lĩnh vực ứng dụng phát hiện khuôn mặt

Bài toán nhận dạng mặt người có thể áp dụng rộng rãi trong nhiều ứng dụng thực tế khác nhau. Đó chính là lý do mà bài toán này hấp dẫn rất nhiều nhóm nghiên cứu trong thời gian dài. Các ứng dụng liên quan đến nhận dạng mặt người có thể kể như:

- Hệ thống phát hiện tội phạm: camera được đặt tại một số điểm công cộng như: siêu thị, nhà sách, trạm xe buýt, sân bay, v.v... Khi phát hiện được sự xuất hiện của các đối tượng là tội phạm, hệ thống sẽ gửi thông điệp về cho trung tâm xử lý. - Hệ thống theo dõi nhân sự trong một đơn vị: giám sát giờ ra vào của từng nhân viên và chấm công.
- Hệ thống giao tiếp người máy: thay thế việc tương tác giữa người và máy theo những cách truyền thống như: bàn phím, chuột, v.v... Thay vào đó là sử dụng các giao tiếp trực quan: biểu cảm khuôn mặt, dấu hiệu, cử chỉ bằng tay (visual input, visual interaction).
- Hệ thống tìm kiếm thông tin trên ảnh, video dựa trên nội dung (chỉ mục theo người). Chẳng hạn như: đài truyền hình Việt Nam (VTV) có một kho dữ liệu video tin tức khá lớn cần tìm kiếm nhanh những đoạn video nào có G. Bush hoặc Bin Laden.
- Các hệ thống bảo mật dựa trên thông tin trắc sinh học: mặt người, vân tay, v.v... thay vì xác nhận mật khẩu, khóa, v.v...

CHƯƠNG 2: PHÁT HIỆN KHUÔN MẶT SỬ DỤNG ĐẶC TRƯNG HAAR

2.1 Đặc trưng của Haar

2.1.1 Đối tượng nhận dạng

Trên ảnh, vùng khuôn mặt là tập hợp các điểm ảnh có những mối quan hệ khác biệt so với các vùng ảnh khác, những mối quan hệ này tạo lên các đặc trưng riêng của khuôn mặt. Tất cả khuôn mặt người đều có chung những đặc điểm sau khi đã chuyển qua ảnh xám, ví dụ như:

- Vùng hai mắt sẽ tối hơn vùng má và vùng cằm, tức mức xám của vùng này cao hơn vượt trội so với hai vùng còn lại.
- Vùng giữa sống mũi cũng tối hơn vùng hai bên mũi.
- ...

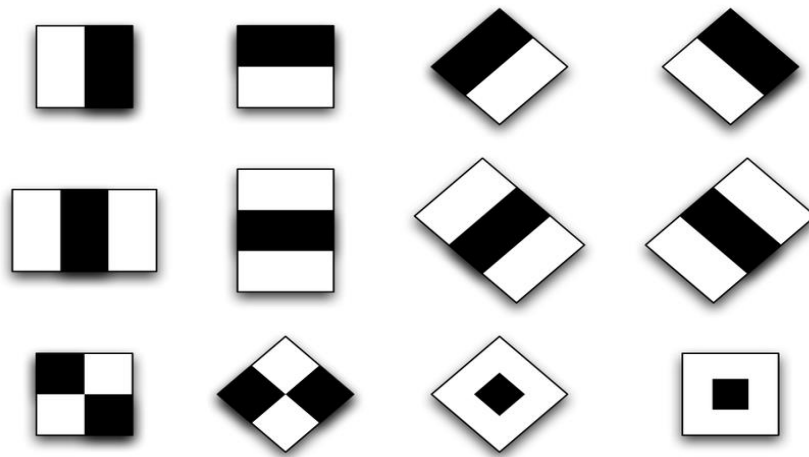
Và còn rất nhiều những đặc điểm khác của khuôn mặt và các đặc trưng Haar like dựa vào các đặc điểm này để nhận dạng.

Về tổng quát, các đặc trưng Haar like không chỉ được sử dụng để nhận dạng khuôn mặt mà có thể dùng để nhận dạng bất kì một đối tượng nào trên ảnh(thân người, tay, chân, ô tô, đồ vật,...). Bởi vì cũng giống như khuôn mặt, mỗi đối tượng có những đặc điểm riêng biệt đặc trưng bởi các vùng điểm ảnh, công việc của đặc trưng Haar-like là tính toán các giá trị tương quan giữa các vùng ảnh đó.

Đây chỉ là bước ban đầu về cách áp dụng đặc trưng để phân loại cửa sổ con, chi tiết cách phân loại của bộ Adaboost và Cascade of Classifiers sẽ trình bày ở mục sau.

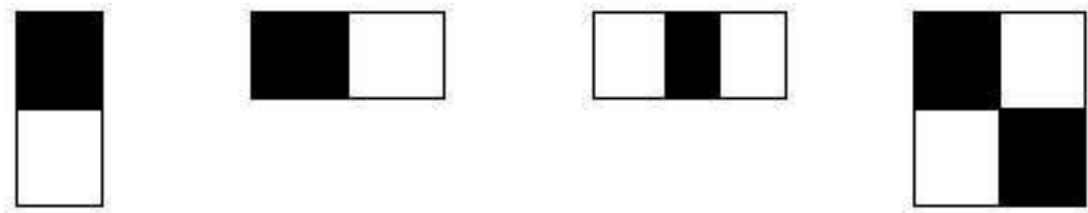
2.1.2 Đặc trưng

Các đặc trưng Haar-Like là những hình chữ nhật được phân thành các vùng khác nhau như hình:



Hình 2-1

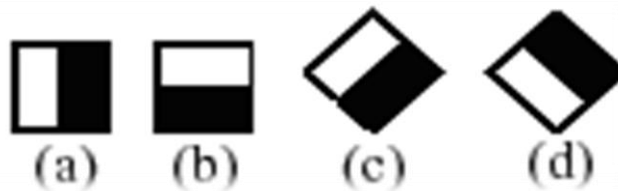
Đặc trưng do Viola và Jones công bố gồm 4 đặc trưng cơ bản để xác định khuôn mặt người. Mỗi đặc trưng Haar-Like là sự kết hợp của hai hay ba hình chữ nhật trắng hay đen như trong hình sau:



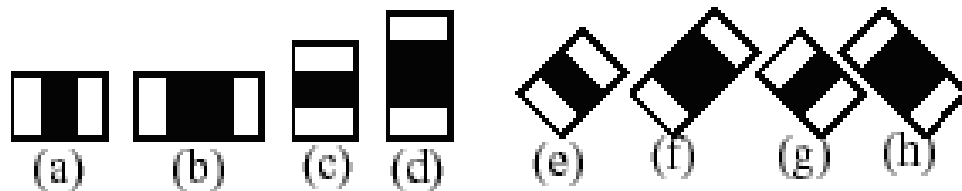
Hình 2-2: 4 đặc trưng Haar-like cơ bản

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

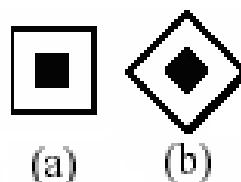
Đặc trưng cạnh(edge feature)



Hình 2-3: Đặc trưng đường(line feature)



Hình 2-4: Đặc trưng xung quanh tâm(center-surround features)



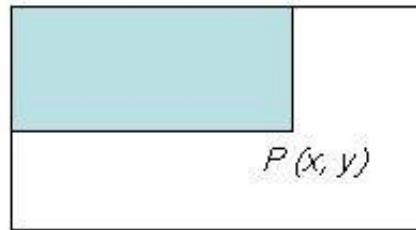
Hình 2-5: Các đặc trưng mở rộng của các đặc trưng Haar-like cơ sở

Dùng các đặc trưng trên, ta có thể tính được các giá trị của đặc trưng Haar-Like là sự chênh lệch giữa tổng của các pixel của vùng đen và vùng trắng như trong công thức sau:

$$f(x) = \text{Tổng}_{\text{vùng đen}}(\text{các mức xám của pixel}) - \text{Tổng}_{\text{vùng trắng}}(\text{các mức xám của pixel}) \quad (2-1)$$

Viola và Joines đưa ra một khái niệm gọi là Integral Image, là một mảng 2 chiều với kích thước bằng với kích thước của ảnh cần tính 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.



Hình 2-6: Công thức tính Intergral Image

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

(2-2)

Sau khi tính được Integral Image, việc tính tổng các giá trị mức xám của một vùng bất kỳ nào đó trên ảnh thực hiện rất đơn giản theo cách sau:

Giả sử ta cần tính tổng giá trị mức xám của vùng D như hình dưới, ta có thể tính được như sau:

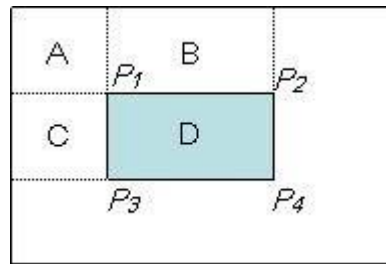
$$D = A + B + C + D - (A+B) - (A+C) + A$$

(2-3)

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:

$$D = \underbrace{(x_4, y_4)}_{A + B + C + D} - \underbrace{(x_2, y_2)}_{(A+B)} - \underbrace{(x_3, y_3)}_{(A + C)} + \underbrace{(x_1, y_1)}_A$$

(2-4)



Hình 2-7: Ví dụ cách tính nhanh các giá trị mức xám của vùng D trên ảnh

Tiếp theo, để chọn các đặc trưng Haar-like dùng cho việc thiết lập ngưỡng, Viola và Jones sử dụng một phương pháp máy học được gọi là AdaBoost. AdaBoost sẽ kết hợp các bộ phân loại yếu để tạo thành một bộ phân loại mạnh. Với bộ phân loại yếu chỉ cho ra câu trả lời chính xác chỉ hơn viện đoán một cách ngẫu nhiên một chút, còn bộ phân loại mạnh có thể đưa ra câu trả lời chính xác trên 60%.

2.2 Bộ phân lớp AdaBoost

2.2.1 Giới thiệu

Bài toán dò tìm khuôn mặt nhanh trên ảnh là bài toán quan trọng vì quá trình nhận dạng đối tượng sẽ thiếu chính xác nếu như thiếu bước dò tìm và định vị được đối tượng. Bài toán dò tìm khuôn mặt nhanh có ý nghĩa rất quan trọng trong việc nhận dạng, theo vết các đối tượng chuyển động trong các đoạn video hay camera.

2.2.2 Các hướng tiếp cận dò tìm khuôn mặt nhanh

- Hướng dò tìm khuôn mặt trên ảnh màu dựa trên sự phân tích màu sắc của vùng da. Mặc dù việc xử lý khá nhanh nhưng hướng này có giới hạn chỉ xử lý trên ảnh màu và thường nhạy cảm với ánh sáng, thường chỉ sử dụng làm các bước tiền xử lý cho các hướng khác.
- Hướng dò tìm khuôn mặt dựa trên đặc trưng bất biến của khuôn mặt người. Thành công nhất trong dò tìm khuôn mặt người trong thời gian thực là phương pháp ASM (Active shape Models).

- Hướng dò tìm khuôn mặt dựa trên thông tin hình ảnh gồm mạng nơ-ron, các hướng thống kê (SVM, AdaBoost....). Phương pháp SVM và mạng nơ-ron cũng đạt được kết quả cao trong thời gian khá nhanh song cũng chỉ vài ảnh trong một giây nên khó có thể áp dụng trong việc nhận diện thời gian thực. Riêng phương pháp AdaBoost cho kết quả khả quan nhất vì có thể xử lý đến khoảng 15-20 khung hình trong một giây.

2.2.3 Thuật toán AdaBoost

Xét bài toán hai lớp, mẫu huấn luyện bao gồm M bộ (x_i, y_i) đã được gán nhãn, với $i \in \{1, 2, \dots, M\}$ trong đó $y_i \in \{+1, -1\}$ là nhãn và $x_i \in \mathbb{R}^n$ là các mẫu huấn luyện. Trong AdaBoost, một bộ phân loại mạnh hơn được xây dựng dựa trên sự kết hợp tuyến tính giữa M bộ phân loại yếu hơn:

$$\mathbf{H}_M(x) = \sum_{m=1}^M \mathbf{h}_m(x) \quad (2-5)$$

Các bộ phân loại yếu hơn có thể mang các giá trị thực, $\mathbf{h}_m(x) \in \mathbb{R}$. Phân loại của x được quyết định bằng hàm $H(x) = \text{sign}[\mathbf{H}_M(x)]$, trong đó độ lớn $|\mathbf{H}_M(x)|$ cho ta độ tin cậy. Mỗi mẫu được kết hợp với một trọng số. Trong quá trình học, các trọng số sẽ được cập nhật động nhấn mạnh các phân loại mạnh trước đó bị phân loại sai. Tuy nhiên, quá trình cập nhật trọng số chỉ cần thiết đối với thuật toán AdaBoost trước đây. Đối với các thuật toán AdaBoost cải tiến gần đây, quá trình này có thể được thay thế bằng một hàm tối ưu hóa. Lỗi xảy ra khi $H(x) \neq y$ hay $y\mathbf{H}_M(x) < 0$. Lỗi của mẫu (x, y) qua hàm $h(x) \in \mathbb{R}$ trên tập các mẫu huấn luyện được định nghĩa là $yh(x)$. Lỗi có thể được xem là số đo độ tin cậy của giá trị đoán trước của h . Lỗi phân lớp của \mathbf{H}_M có biên trên là:

$$J(\mathbf{H}_m) = \sum e^{-y_i \mathbf{H}_m(x_i)} \quad (2-6)$$

Thuật toán AdaBoost xây dựng hàm $h(x)$ bằng cách giảm tối đa (2-6).

Cho

$$\mathbf{H}_{M-1}(x) = \sum_{m=1}^{M-1} h_m(x) \quad (2-7)$$

$H_M(x)$ tốt nhất cho phân loại mạnh $\mathbf{H}_M(x) = \mathbf{H}_{M-1}(x) + h_m(x)$

Là hàm dẫn tới giá trị nhỏ nhất:

$$H_m = \arg \min J(H(x) + h(x)) \quad (2-8)$$

và hàm có giá trị nhỏ nhất được chứng minh là:

$$h_M(x) = \frac{1}{2} \log \frac{P(y=+1|x, \omega^{(M-1)})}{P(y=-1|x, \omega^{(M-1)})} \quad (2-9)$$

Với $\omega^{(M-1)}$ trọng lượng tại thời điểm M .

Dùng công thức $P(y|x, \omega) = P(x|y, \omega) P(y)$ và cho

$$L_M(x) = \frac{1}{2} \log \frac{P(x|y=+1, \omega)}{P(x|y=-1, \omega)} \quad (2-10)$$

$$T = \frac{1}{2} \log \left[\frac{P(y=+1)}{P(y=-1)} \right] \quad (2-11)$$

Chúng ta có được $H_m(x) = L_m(x) - T$. L_m được học ra từ các mẫu của cả hai phân lớp. Ngưỡng T được xác định bằng tỉ lệ log của các xác suất trước đó.

Ta có một phương pháp để tính phương trình (2-11), ứng dụng khi học các bộ phân lớp tối ưu. Vì rút ra một bộ phân loại yếu trong một miền không gian nhiều chiều là công việc quan trọng, xin được đưa ra sau đây một mô hình thống kê học theo từng giai đoạn dựa trên vài đặc điểm vô hướng. Một đặc điểm vô hướng j của x được tính bằng một phép biến đổi từ không gian dữ liệu n chiều thành đường thẳng thực $Z_j(x) \in Z$. Một đặc điểm có thể là hệ số, hay nói trong xử lý ảnh là phép biến đổi vi ba tín hiệu. Nếu phương pháp tìm kiếm ước lượng được sử dụng như phép biến đổi $Z_j(x)$ đơn giản được xem là tọa độ thứ j của x . Một danh sách K đặc điểm ứng cử viên có thể được tạo $Z = \{Z_1(x), \dots, Z_k(x)\}$. Trong phần sau, chúng ta sử dụng $z(m)$ để biểu diễn cho đặc điểm được chọn trong giai đoạn m , và $Z_k(x)$ là đặc điểm được tính toán từ x sử dụng phép biến đổi thứ k .

Giả sử Z là một tập rất hoàn chỉnh, tập các phân lớp yếu có thể có cho bài toán phân lớp yếu tối ưu có thể được lập như sau: Trước tiên, tại giai đoạn M , khi $M-1$ đặc điểm của $Z_{(1)}, Z_{(2)}, \dots, Z_{(M-1)}$ đã được chọn và trọng lượng số cho là $\omega^{(M-1)}$, chúng ta xấp xỉ $p(x|y, \omega^{(M-1)})$ bằng cách dùng phân bố của M đặc điểm:

$$p(x|y, \omega^{(M-1)}) \approx p(Z_{(1)}, Z_{(2)}, \dots, Z_{(M-1)}, Z_k, |y, \omega^{(M-1)}) \quad (2-12)$$

$$= p(Z_{(1)}|y, \omega^{(M-1)}) p(Z_{(2)}|y, Z_{(1)}, \omega^{(M-1)}) \dots$$

$$p(Z_{(M-1)}|y, Z_{(1)}, Z_{(2)}, \dots, Z_{(M-2)}, \omega^{(M-1)})$$

$$p(Z_k, |y, Z_{(1)}, Z_{(2)}, \dots, Z_{(M-1)}, \omega^{(M-1)}) \quad (2-13)$$

Bởi vì Z là tập rất hoàn chỉnh, phép xấp xỉ vẫn tốt đối với tập M đủ lớn khi M đặc điểm được chọn thích hợp.

Ghi chú: $p(\mathbf{Z}_m|y, \mathbf{Z}_{(1)}, \mathbf{Z}_{(2)}, \dots, \mathbf{Z}_{(m-1)})$ thực ra là $p(\mathbf{Z}_m|y, \boldsymbol{\omega}^{(M-1)})$ bởi vì $\boldsymbol{\omega}^{(M)}$ chứa thông tin về toàn bộ quá trình tạo ω và bao gồm các thành phần lệ thuộc trên $\mathbf{Z}_{(1)}, \mathbf{Z}_{(2)}, \dots, \mathbf{Z}_{(m-1)}$. Vì vậy, chúng ta có:

$$p(x|y, \boldsymbol{\omega}^{(M-1)}) \approx p(\mathbf{Z}_{(1)}|y, \boldsymbol{\omega}^{(0)}) p(\mathbf{Z}_{(2)}|y, \boldsymbol{\omega}^{(1)}) \dots$$

$$p(\mathbf{Z}_{(M-1)}|y, \boldsymbol{\omega}^{(M-2)}) p(\mathbf{Z}_k|y, \boldsymbol{\omega}^{(M-1)}) \quad (2-14)$$

Mật độ xác suất $p(\mathbf{Z}_k|y, \boldsymbol{\omega}^{(M-1)})$ cho phân lớp dương $y = +1$ và phân lớp âm $y = -1$ có thể phỏng đoán được từ histogram tính được qua đánh giá công nhận trọng số của các ví dụ huấn luyện sử dụng các trọng số $\boldsymbol{\omega}^{(M-1)}$.

Cho

$$L_k^{(M)}(\mathbf{x}) \frac{P(\mathbf{Z}_k|y=+1, \boldsymbol{\omega}^{(M-1)})}{P(\mathbf{Z}_k|y=-1, \boldsymbol{\omega}^{(M-1)})} \quad \text{và} \quad h_k^{(M)} = \frac{1}{2} L_k^{(M)}(\mathbf{x}) - T \quad (2-15)$$

chúng ta rút ra được tập hợp các phân lớp yếu hơn như sau:

$$\tau^{(M)}(\mathbf{x}) = h_k^{(M)}(\mathbf{x}) | \forall k \quad (2-16)$$

Thuật toán Adaboost

Bước 0. Đầu vào

1. Tập $Z = \{(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2), \dots, (\mathbf{x}_n, \mathbf{y}_n)\}$ với:

$$N = a + b;$$

a là số mẫu thuộc phân lớp $\mathbf{y}_i = +1$

b là số mẫu thuộc phân lớp $\mathbf{y}_i = -1$

2. Số lớp yếu tối đa M_{\max} được kết hợp

Bước 1: Khởi tạo

$$\omega_i^{(0)} = \frac{1}{2a} \quad \text{với mẫu thuộc phân lớp } y_1 = +1$$

$$\omega_i^{(0)} = \frac{1}{2a} \quad \text{với mẫu thuộc phân lớp } y_1 = -1$$

$$M = 0$$

Bước 2: Suy diễn tiến

While $M < M_{\max}$

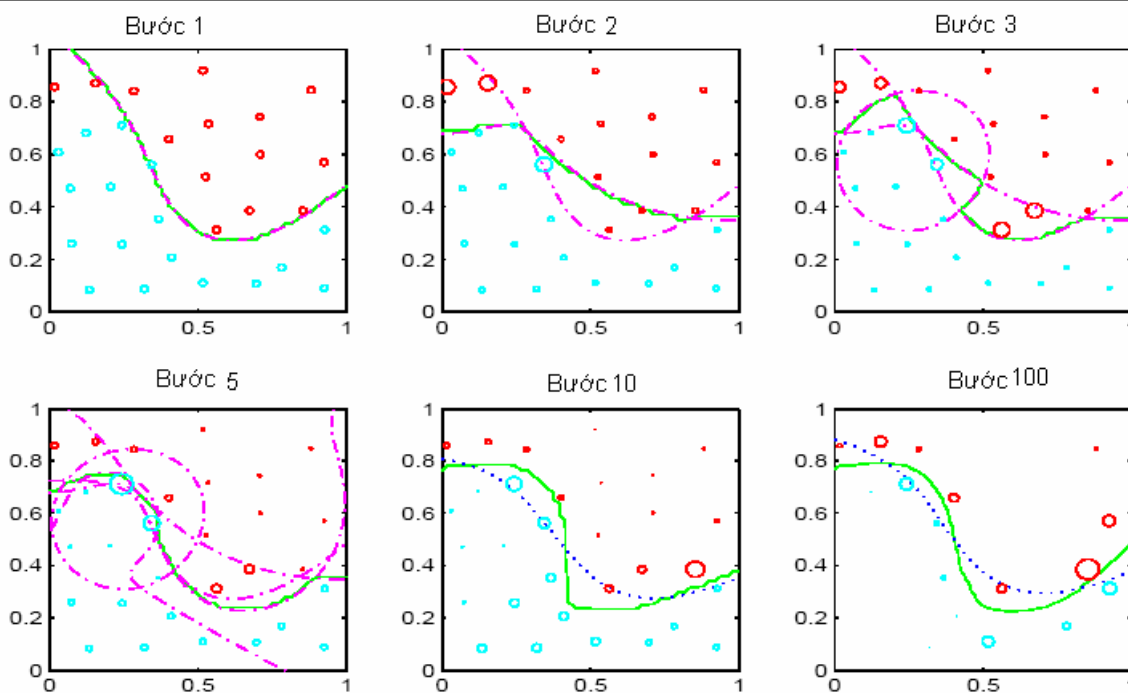
1) $M \leftarrow M+1$

2) Chọn H_m theo biểu thức (2-8)

3) Cập nhật $\omega_i^{(m)} \leftarrow \exp -y_i H_m x_{(i)}$ và chuẩn hóa $\omega_i^{(m)}$
để $\sum \omega_i^{(m)} = 1$

Bước 3: Đầu ra

$$H(x) = \text{sign} \sum_{M=1}^M h_m(x)$$



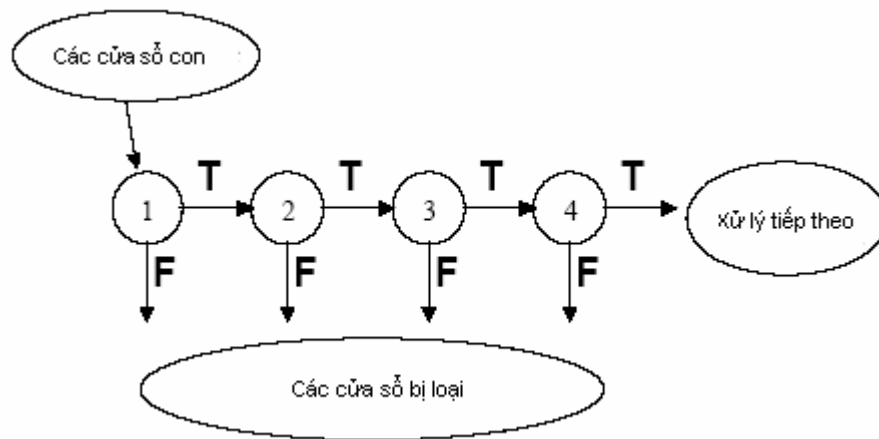
Hình 2-8: Ví dụ minh họa cho thuật toán AdaBoost

2.2.4 Bộ phân lớp AdaBoost

Với một bộ dò tìm c do phương pháp AdaBoost huấn luyện được, ta có thể dò tìm với một độ chính xác nhất định và một tốc độ nhất định. Nếu như cần phải chính xác cao thì bộ dò tìm phải bao gồm nhiều đặc trưng, điều đó kéo theo tốc độ dò tìm sẽ giảm.

Nếu sử dụng bộ dò tìm kết hợp $F = \{C_i\}$ với nhiều bộ dò tìm cơ bản f_i khác nhau cũng rơi vào tình trạng tương tự. Để có được độ chính xác cao, hoặc cần phải có số lượng lớn các bộ dò tìm, hoặc mỗi bộ dò tìm cần phải có nhiều đặc trưng, hoặc cả hai. Do đó cũng kéo theo tốc độ sẽ giảm. Một hướng khắc phục nhược điểm này là sử dụng bộ dò tìm phân tầng $T = \{t_i\}$. Bộ dò tìm phân tầng bao gồm nhiều tầng, mỗi tầng $t_i = \{c_i\}$ là một bộ dò tìm kết hợp với số lượng các bộ dò tìm khác nhau nên có tốc độ và độ chính xác khác nhau. Khi dò tìm tất cả các khuôn mặt trong ảnh, tất cả các cửa sổ con $W_0 = \{w_{i,j,s}\}$ với các kích thước s khác nhau tại các tọa độ (i,j) sẽ được kiểm tra xem có phải là mặt người hay không. Qua mỗi tầng t_i , $W_i = t_i(W_{i-1})$

trong đó $|W_i| \ll |W_{i-1}|$ do qua mỗi tầng, các cửa sổ không phải ứng viên sẽ bị loại sớm. Điều này cho phép chúng ta xây dựng các tầng sao cho, càng về sau độ phức tạp (số lượng các bộ dò tìm cơ bản với các đặc trưng) càng lớn trong khi các tầng càng thấp thì độ phức tạp càng đơn giản và phải loại được nhiều ứng viên càng tốt nhưng tỷ lệ loại sai phải thấp.



Hình 2-9: Minh họa bộ đồ tìm phân tầng

Xét mỗi tầng $t_k = \{C_i\}$ ta có tỷ lệ loại sai của t_k được tính như sau:

$$F = \prod_{i=1}^k f_i \quad (2-17)$$

Trong đó f_i chính là tỷ lệ loại sai ứng với bộ dò tìm c_i và K chính là số bộ dò tìm của tầng t_k . Tương tự, độ chính xác của tầng t_k được tính như sau :

$$D = \prod_{i=1}^k d_i \quad (2-18)$$

Trong đó di là độ chính xác của ứng với bộ dò tìm ci. Đồng thời cũng với cách tính này, ta có thể tính được độ chính xác của toàn bộ các tầng $T=\{t_i\}$ là:

$$G = \prod_{i=1}^{|T|} D_i = \prod_{i=1}^{|T|} \cdot \prod_{j=1}^{|t_i|} C_j^i \quad (2-19)$$

Vậy khi cho trước một tỷ lệ loại sai D và độ chính xác là F, ta có thể huấn luyện tầng bộ phân loại t sao cho t có tỷ lệ loại sai là D và độ chính xác là F. Và lặp lại quá trình huấn luyện tầng ta được bộ huấn luyện gồm nhiều tầng với độ chính xác G hoặc số tầng n như mong muốn. Dưới đây là thuật toán huấn luyện một tầng với tỷ lệ loại sai f và độ chính xác d cho trước:

Bước 0. Đầu vào:

f (tỷ lệ nhận sai mẫu dương tối đa chấp nhận được)

d (tỷ lệ nhận đúng tối thiểu trong lớp)

Ftarget (tỷ lệ nhận sai mẫu dương)

P=tập mẫu dương

N=tập mẫu âm

Bước 1. Khởi tạo:

F0=1.0

D0 = 1.0

i = 0

Bước 2:

Trong khi (Fi > Ftarget)

i ← i + 1

$$n_i = 0; F_i = F_i - 1$$

Trong khi ($F_i > f \times F_i - 1$)

$$n_i = n_i + 1$$

- Sử dụng P và N để huấn luyện một phân loại H vs n_i đặc trưng, dùng AdaBoost
- Thêm bộ phân loại hiện thời vào C
- Tính F_i và D_i cho bộ phân loại C hiện thời lên tập hợp lẹ
- Giảm ngưỡng cho lớp thứ I cho đến khi bộ phân loại C hiện thời đạt tỉ lệ dò tìm tối thiểu là $d \times D_{i-1}$ (điều này cũng ảnh hưởng tới F_i)
- $N \leftarrow \emptyset$
- Nếu $F_i > F_{\text{target}}$ thì định giá bộ dò tìm C hiện thời trên tập ảnh không phải mặt người và đưa các mẫu dò tìm bị lỗi vào tập N

Bước 3. Đầu ra: bộ dò tìm đa tầng C

2.3 Dò tìm khuôn mặt

Việc dò tìm khuôn mặt trong ảnh qua bộ phân tầng đã huấn luyện gặp một vấn đề là số sửa số với các kích thước khác nhau quá lớn. Để khắc phục vấn đề này, phương pháp dò theo kiến trúc tháp được áp dụng như sau: xét ảnh s kích thước ($w \times h$), $\text{step} = 0$, hệ số co scale 1.2

Lặp trong khi kích thước ($w \times h$) còn lớn hơn cửa sổ ảnh mặt người huấn luyện ($w_0 \times h_0$):

- Duyệt toàn bộ các vị trí (x, y) cửa sổ với kích thước ($w_0 \times h_0$), với mỗi đơn vị trí tiến hành:
- Áp dụng bộ dò tìm phân tầng để xác định có phải mặt người hay không

- Nếu là mặt người tại vị trí (x,y) thì thực tế mặt người tại vị trí $(x * scale^{step}, y * scale^{step})$ và kích cỡ cửa sổ là $(w_0 * scale^{step}, h_0 * scale^{step})$
- Gán $w_1 = W / scale$ và $h_1 = h / scale$
- Thu nhỏ ảnh từ kích thước $(w \times h)$ đến $(w_1 \times h_1)$
- Gán $w = w_1$ và $h = h_1$
- $step = step + 1$

Nhận xét:

Hệ số co scale quyết định độ mịn của các cửa sổ dò tìm, nếu như scale càng nhỏ (≥ 1) thì càng có nhiều cửa sổ dò tìm nên càng chính xác hơn.

Áp dụng thuật toán dò theo kiến trúc tháp như trên ta có thể dò tìm tất cả các khuôn mặt ở tất cả các vị trí, song kích thước dò tìm ở mỗi bước như sau:

Bước 1: Kích thước từ (w_0, h_0) đến $(w_0 * scale, h_0 * scale)$

Bước 2: Kích thước từ $(w_0 * scale, h_0 * scale)$ đến $(w_0 * scale^2, h_0 * scale^2)$

.....

Bước n: Kích thước từ $(w_0 * scale^{n-1}, h_0 * scale^{n-1})$ đến $(w_0 * scale^n, h_0 * scale^n)$

2.4 Nhận xét

2.4.1 Ưu điểm

- Phương pháp cho độ chính xác tương đối cao (trên 90%), tốc độ dò tìm khuôn mặt khá nhanh, thích hợp để dò tìm khuôn mặt trong thời gian thực, trong video.
- Thích hợp với việc huấn luyện dữ liệu bị nhiễu.
- Phương pháp trích chọn đặc trưng thực hiện khá nhanh.

2.4.2 Nhược điểm

- Thuật toán huấn luyện khuôn mặt thực hiện chậm do có rất nhiều bộ phân loại yếu.
- Chỉ dò tìm được các khuôn mặt nhìn thẳng và góc quay nhỏ
- Không ít các tính chất của AdaBoost mang tính chất nhận định và chưa được chứng minh chặt chẽ.

CHƯƠNG 3: THỰC NGHIỆM PHÁT HIỆN KHUÔN MẶT

3.1 Mô tả bài toán

Từ các phương pháp được đưa ra ở phần trước, bài toán phát hiện khuôn mặt được thực hiện theo hai bước chính:

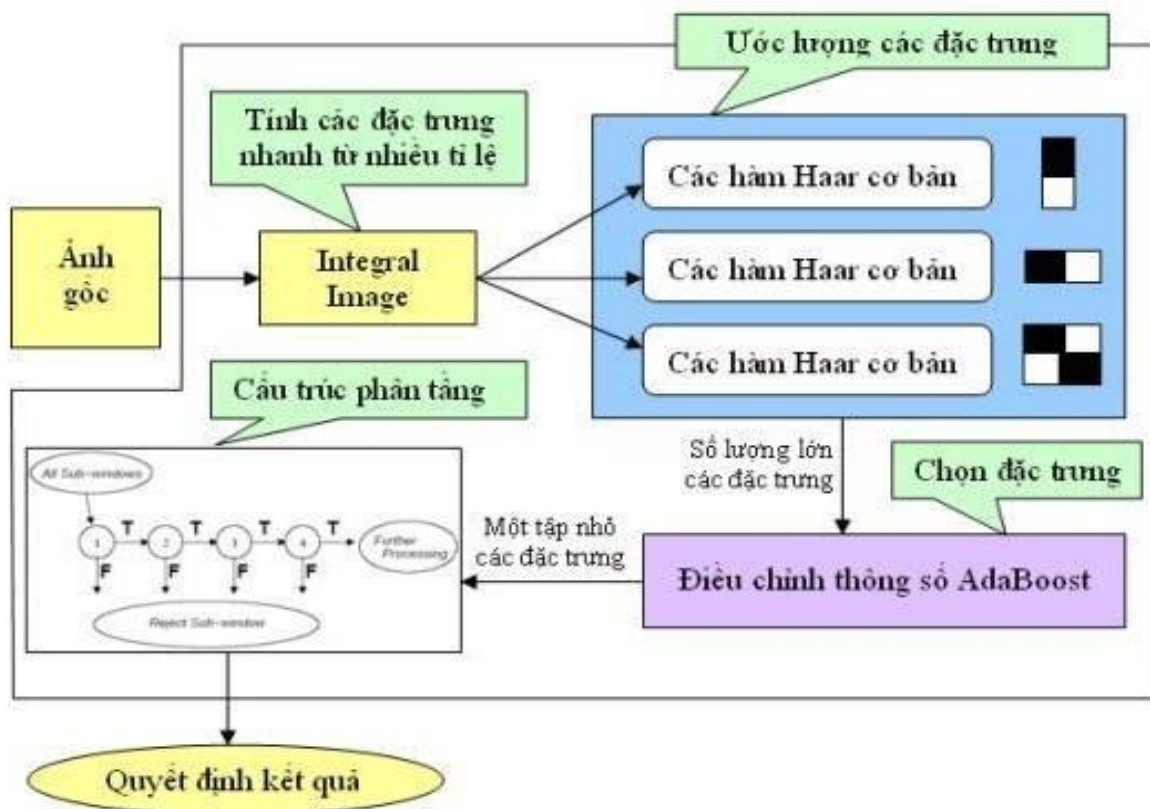
- Do tìm các vùng có thể là khuôn mặt trong ảnh.
- Xác thực khuôn mặt.

Phân tích và lựa chọn công cụ

Đầu vào: Ảnh gốc.

Đầu ra: ảnh đã được phát hiện khuôn mặt.

Để nhận dạng khuôn mặt bằng đặc trưng haar like thì đầu tiên ta phải đưa ảnh về xám hóa, sau đó cho các đặc trưng haarlike chạy khắp bức ảnh, những khu vực so sánh được cho là giống với nhiều đặc trưng haar like nhất sẽ được đánh dấu lại.



Hình 3-1: Hệ thống phát hiện khuôn mặt người

Từ ảnh gốc ban đầu ta sẽ tính Integral image, là mảng hai chiều với phần tử (x,y) sẽ được tính bằng tổng của các phần tử (x',y') với $x' < x$ và $y' < y$, làm như vậy để tính nhanh tổng của các giá trị mức xám của một vùng hình chữ nhật bất kỳ trên ảnh gốc. Các vùng ảnh con này sẽ được đưa qua các hàm Haar cơ bản để ước lượng đặc trưng, kết quả ước lượng sẽ được đưa qua bộ điều chỉnh Adaboost để loại bỏ nhanh các đặc trưng không có khả năng là các đặc trưng của khuôn mặt người. Chỉ có một tập nhỏ các đặc trưng mà bộ điều chỉnh Adaboost cho là có khả năng là đặc trưng của khuôn mặt người mới được chuyển sang cho bộ quyết định kết quả (là tập các bộ phân loại yếu). Bộ quyết định sẽ tổng hợp kết quả là khuôn mặt người nếu kết quả của các bộ phân loại yếu trả về là khuôn mặt người.

Mỗi bộ phân loại yếu sẽ quyết định kết quả cho một đặc trưng Haar-like, được xác định ngưỡng đủ nhỏ sao cho có thể vượt được tất cả các bộ dữ liệu mẫu trong tập dữ liệu huấn luyện (số lượng ảnh khuôn mặt trong tập huấn luyện).

3.2 Phát hiện khuôn mặt sử dụng hàm OpenCV

3.2.1 OpenCV là gì

OpenCV viết tắt của từ Open Source Computer Vision Library.

- OpenCV là một thư viện mã nguồn mở phục vụ cho việc nghiên cứu hay phát triển về thị giác máy tính.
- Tối ưu hóa và xử lý ứng dụng trong thời gian thực.
- Giúp cho việc xây dựng các ứng dụng xử lý ảnh, thị giác máy tính ... một cách nhanh hơn.
- OpenCV có hơn 500 hàm khác nhau, được chia làm nhiều phần phục vụ các công việc như: xử lý ảnh, an ninh, camera quan sát, nhận diện, robot...

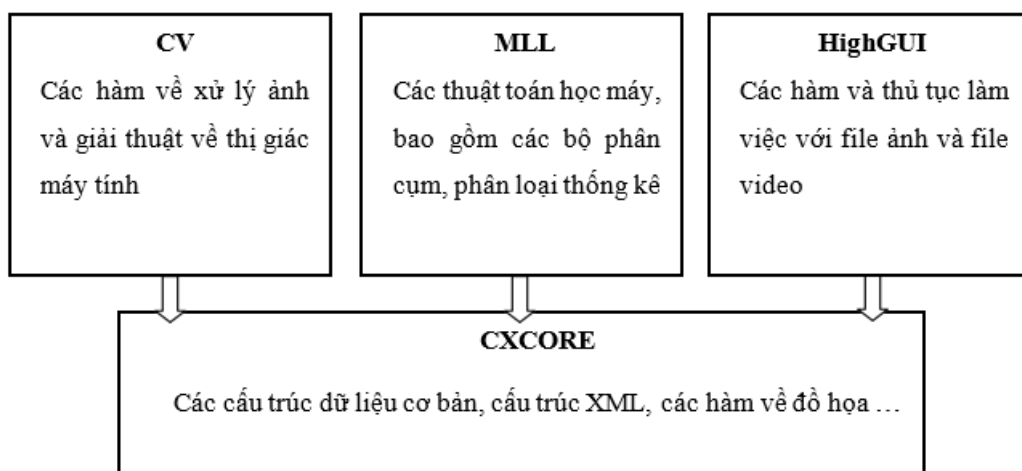
Thư viện được viết bằng ngôn ngữ C và C++ có thể chạy trên các hệ điều hành Linux, Window và MacOSX. OpenCV được thiết kế để nâng cao hiệu suất tính toán và nhấn mạnh đến hệ thống thời gian thực. OpenCV đưa ra một hệ thống đơn giản, dễ sử dụng giúp mọi người nhanh chóng xây dựng các ứng dụng trong thị giác máy, kể cả các hệ thống kiểm tra trong nhà máy, bức ảnh trong lĩnh vực y học, bảo mật, robot học...

Nó chứa các lập trình xử lý ảnh đơn giản, kể cả thực thi các hàm bậc cao như dò tìm khuôn mặt, theo dõi khuôn mặt, nhận dạng khuôn mặt...

OpenCV được giới thiệu vào tháng 1/1999, OpenCV đã được sử dụng trong rất nhiều ứng dụng, các sản phẩm và các nghiên cứu như: trong lĩnh vực hàng không, sử dụng giảm nhiễu trong y học, phân tích đối tượng, an ninh, hệ thống dò tìm, theo dõi tự động và hệ thống bảo mật...., ngoài ra nó còn được sử dụng trong nhận dạng âm thanh. OpenCV còn là một chìa khóa quan trọng trong các robot sử dụng thị giác như Stanford, Asimo.

3.2.2 Cấu trúc OpenCV

Cấu trúc của opencv được chia làm 5 phần chính, 4 trong số đó được chia ra như trong hình sau:



Hình 3-2: Cấu trúc cơ bản của OpenCV

CV (computer vision) là thành phần chứa những xử lý ảnh cơ sở và thuật toán thị giác máy tính mức cao.

MLL (machine learning library) là thư viện machine learning, cái này bao gồm rất nhiều lớp thống kê và gộp công cụ xử lý.

HighGUI chứa các thủ tục vào ra và các hàm dùng cho việc lưu trữ và tải ảnh video.

CXCore chứa cấu trúc và nội dung dữ liệu cơ sở.

3.2.3 Phát hiện khuôn mặt với các hàm trong OpenCV

Từ các phương pháp được đưa ra ở phần trước, bài toán phát hiện khuôn mặt được thực hiện theo các bước:

- Huấn luyện bộ tập mẫu
- Xác định khuôn mặt

Ở bước huấn luyện bộ tập mẫu. Trong khi cài đặt OpenCV 3.4.2 đã có sẵn thư mục bộ huấn luyện có sẵn. Ta sẽ sử dụng luôn tài nguyên này.

Ở bước xác định khuôn mặt sẽ chia làm 3 bước:

- Tiền xử lí
- Dò tìm khuôn mặt
- Hậu xử lí

3.2.3.1 Huấn luyện bộ tập mẫu

Ta sử dụng luôn bộ huấn luyện có sẵn trong thư mục của opencv

```
CascadeClassifier faceDetection;  
if (!faceDetection.load("D:\\opencv\\sources\\data\\haarcascades\\haarcascade_frontalface_default.xml"));
```

Hình 3-3: Trong thư mục của OpenCV đã có các bộ huấn luyện có sẵn.

3.2.3.2 Xác định khuôn mặt

Bước 1: Tiền xử lí

Một lưu ý nhỏ là phương pháp HA thực hiện trên ảnh xám (gray image). Mỗi điểm ảnh (pixel) sẽ có giá trị mức xám từ 0 đến 255 (không gian màu 8 bit). Như vậy phương pháp HA sẽ không khai thác những đặc điểm về màu sắc khuôn mặt để nhận dạng song vẫn rất hiệu quả. Ảnh màu sẽ được chuyển về ảnh xám để nhận dạng, việc chuyển đổi này khá đơn giản, thực hiện bằng một hàm chuyển đổi và sử dụng chỉ với một câu lệnh trong Opencv nên báo cáo chưa đề cập tới. Sau khi chuyển thành ảnh xám, ảnh lại tiếp tục được chuyển thành “ảnh tích hợp” và trong bước đầu tiên của quá trình nhận dạng, các đặc trưng Haar-like sẽ làm việc trực tiếp trên ảnh tích hợp.



Hình 3-4: Chuyển về ảnh xám

Bước 2: Dò tìm khuôn mặt

Integral Image là mảng 2 chiều với kích thước ảnh cần tính đặc trưng Haar. Bắt đầu từ vị trí trên bên trái đến vị trí dưới, bên phải 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 sẽ đơn giản hơn.

Để phát hiện khuôn mặt, hệ thống sẽ cho một cửa sổ con (sub-window) có kích thước cố định quét lên toàn bộ ảnh đầu vào. Như vậy sẽ có rất nhiều ảnh con ứng với từng cửa sổ con, các đặc trưng Haar-like sẽ được đặt lên các cửa sổ con này để từ đó tính ra giá trị của đặc trưng. Sau đó các giá trị này được bộ huấn luyện xác nhận xem khung hình đó có phải khuôn mặt hay không.

Hình dưới là một ví dụ: khung màu xanh là cửa sổ con, một đặc trưng Haar-like với kích thước và vị trí đặt như trong hình sau:



Hình 3-5: Tầm nhìn của máy tính khi các đặc trưng Haar chạy

Bước 3: Hậu xử lí

Sau khi quét các cửa sổ con khắp bức ảnh. Đặc trưng Haar sẽ so sánh với bộ huấn luyện mẫu và sẽ đánh dấu những nơi có thể là khuôn mặt trong ảnh.

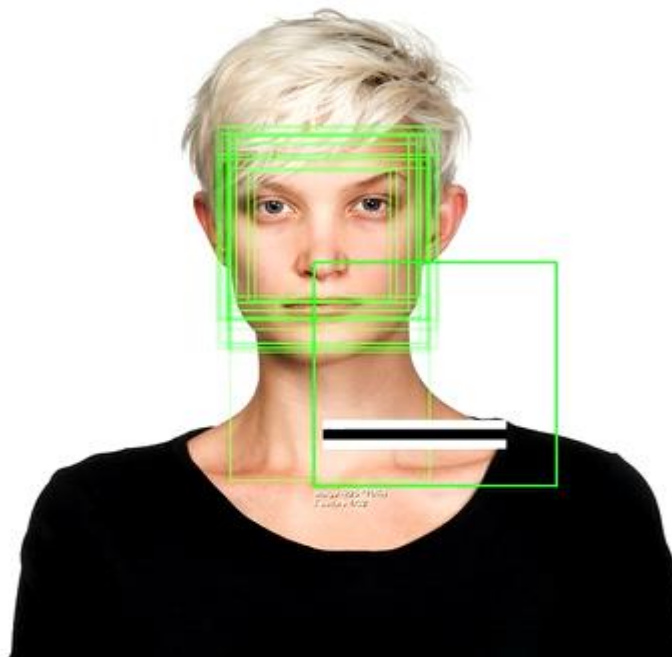
```
faceDetection.detectMultiScale(img, faces);  
  
for (int i = 0; i < faces.size(); i++) {  
    Point pt1(faces[i].x, faces[i].y);  
    Point pt2((faces[i].x + faces[i].height), (faces[i].y + faces[i].width));  
    rectangle(img, pt1, pt2, Scalar(0, 0, 255), 2, 8, 0);  
}
```

Hình 3-6

Trong đó: faces.width/height: Chiều rộng/cao của đặc trưng Haar.

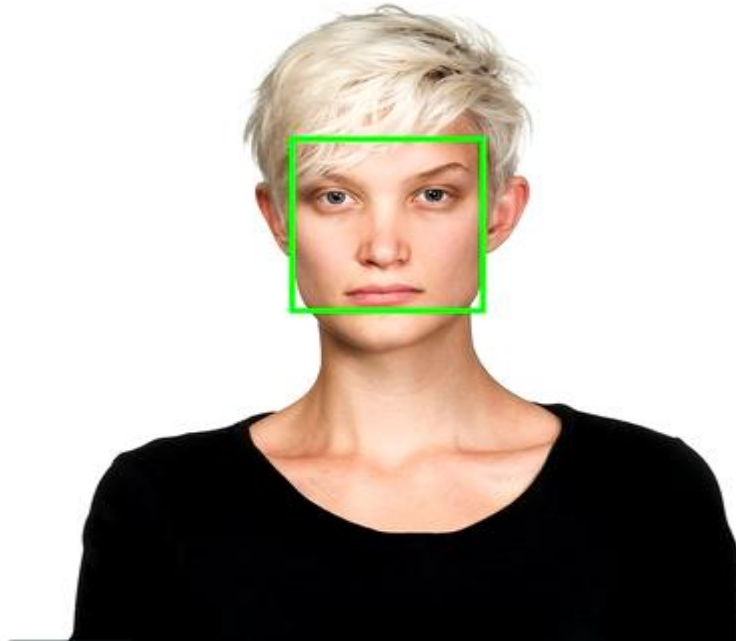
Hàm `detectMultiScale` là phần tìm kiếm khuôn mặt, hàm (hay phương thức) này thuộc lớp `CascadeClassifier` (lớp phục vụ tìm kiếm đối tượng của `Opencv`)

Hàm `detectMultiScale` sau khi tìm kiếm xong sẽ trả về bộ giá trị gồm tọa độ góc của khung chứa khuôn mặt x,y ; chiều dài, rộng của khung w, h . Các giá trị này nằm trong mảng `faces`. Cấu trúc `for...` sẽ duyệt qua toàn bộ các bộ giá trị này, với mỗi bộ giá trị ta dùng hàm `rectangle` để vẽ một hình chữ nhật lên ảnh ban đầu với tọa độ 2 điểm trái trên và phải dưới: $(x,y), (x+w,y+h)$. $(0,0,255)$ là màu sẽ vẽ hình chữ nhật.



Hình 3-7: Các đặc trưng Haar đánh dấu những nơi có thể là khuôn mặt

Sau khi đã đánh dấu hết những nơi có thể là khuôn mặt thì sẽ gộp lại những hình chữ nhật chứa cùng một khuôn mặt để chỉ cho ra một hình chữ nhật cho 1 khuôn mặt.

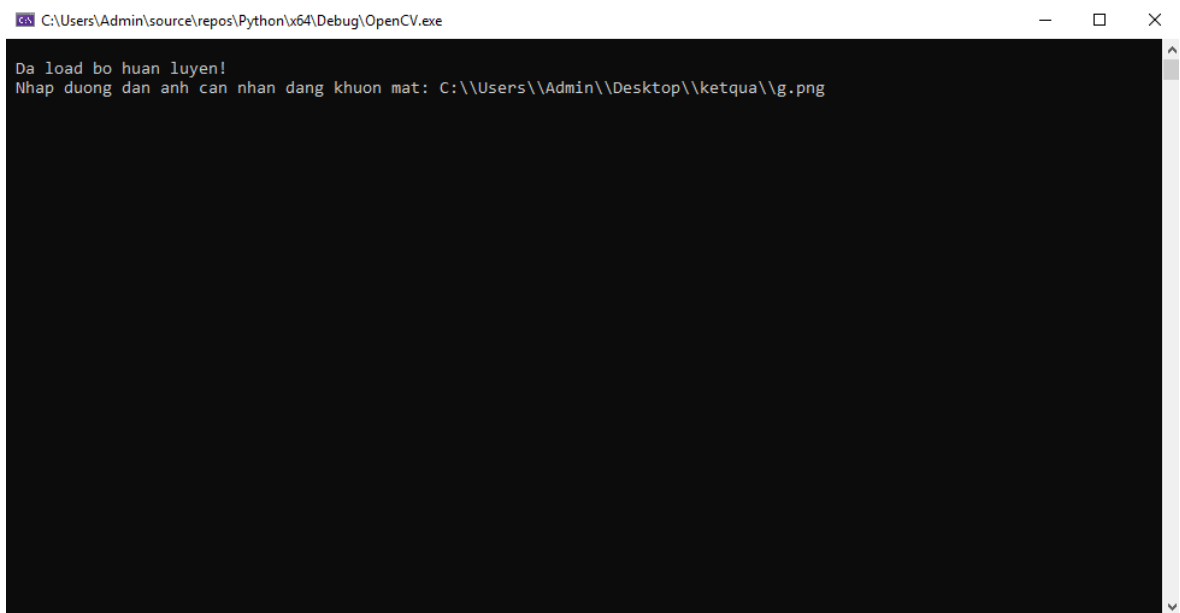


Hình 3-8: Bức ảnh đã được xác định khuôn mặt

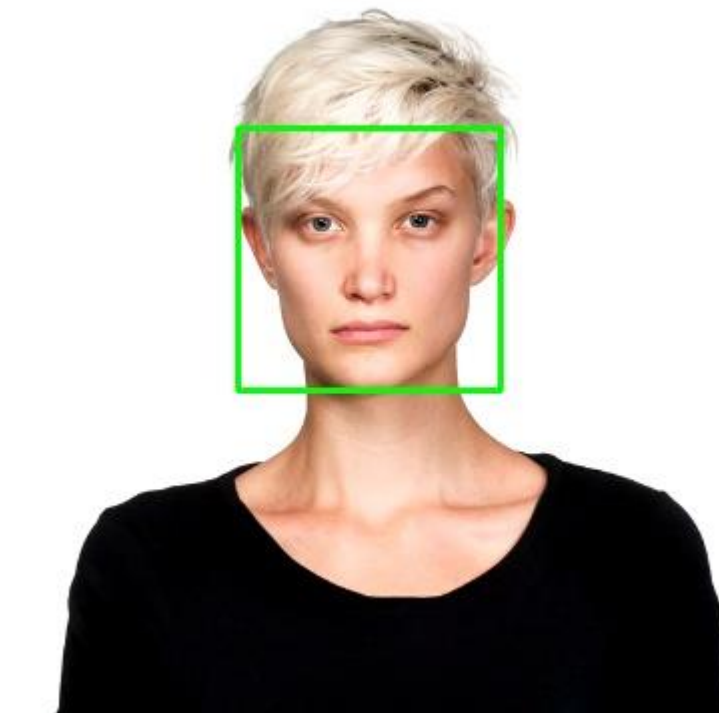
3.3 Một số kết quả chương trình



Hình 3-9: Bức ảnh căn nhận diện khuôn mặt



Hình 3-10: Giao diện chương trình



Hình 3-11: Kết quả

Thử nghiệm với ảnh có nhiều khuôn mặt:



Hình 3-12: Bức ảnh có nhiều khuôn mặt



Hình 3-14: Kết quả (khuôn mặt nghiêng chưa nhận dạng được)

KẾT LUẬN

Qua tìm hiểu bài toán phát hiện mặt người trong ảnh em đã thấy được sự quan trọng của xử lý ảnh cũng như sự phát triển và tầm quan trọng của bài toán phát hiện mặt người trong đời sống.

Trong quá trình xây dựng chương trình em đã tìm hiểu về thư viện mã nguồn mở OpenCV. Qua đó biết cách sử dụng các hàm được xây dựng sẵn trong thư viện OpenCV.

Về chương trình em mới chỉ thực hiện phát hiện mặt người trong ảnh, kết quả đạt được như mong muốn, thời gian phát hiện nhanh đối với mặt người chụp thẳng, chất lượng ảnh tốt, không có vật cản...., còn trong các ảnh như chụp hơi nghiêng hay ảnh có màu sắc tối thì chương trình vẫn chưa phát hiện ra khuôn mặt trong ảnh đó. Em sẽ tìm hiểu thêm và cố gắng khắc phục.

TÀI LIỆU THAM KHẢO

- [1.] <https://123doc.org/document/2903809-bao-cao-do-an-phat-hien-khuon-mat-voi-thuat-toan-adaboost.htm>
- [2.] <https://www.stdio.vn/articles/opencv-voi-python-trong-ung-dung-phat-hien-khuon-mat-trong-buc-anh-617>