

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



ISO 9001:2008

ĐỒ ÁN TỐT NGHIỆP

NGÀNH: CÔNG NGHỆ THÔNG TIN

HẢI PHÒNG - 2013

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

**TÌM HIỂU BÀI TOÁN PHÁT HIỆN TRẠNG THÁI
MẮT CỦA MẶT NGƯỜI TRONG ẢNH**

ĐỒ ÁN TỐT NGHIỆP ĐẠI HỌC HỆ CHÍNH QUY

NGÀNH: CÔNG NGHỆ THÔNG TIN

Sinh viên thực hiện : Nguyễn Thị Hiền

Giáo viên hướng dẫn : PGS. TS. Đỗ Năng Toàn

Mã số sinh viên : 1351010020

HẢI PHÒNG - 2013

MỤC LỤC

DANH MỤC HÌNH VẼ	3
LỜI CẢM ƠN	3
MỞ ĐẦU	4
Chương 1: KHÁI QUÁT VỀ XỬ LÝ ẢNH VÀ PHÁT HIỆN MẮT CỦA MẶT NGƯỜI TRONG ẢNH	6
1.1 Khái quát về xử lý ảnh	6
1.1.1 Xử lý ảnh là gì?	6
1.1.2 Các vấn đề cơ bản trong xử lý ảnh	6
1.1.2.1 Một số khái niệm cơ bản	6
1.1.2.2 Nắn chỉnh biến dạng	7
1.1.2.3 Khử nhiễu	8
1.1.2.4 Chỉnh mức xám	8
1.1.2.5 Trích chọn đặc điểm	8
1.1.2.6 Nhận dạng	9
1.1.2.7 Nén ảnh	11
1.2 Phát hiện trạng thái mắt của mặt người trong ảnh	13
1.2.1 . Giới thiệu	13
1.2.2. Phát hiện mặt người trong ảnh	13
1.2.3. Phát hiện mắt người trong ảnh	14
1.2.4. Phát hiện trạng thái mặt người theo một số phương pháp	15
Chương 2: PHÁT HIỆN MẮT CỦA MẶT NGƯỜI TRONG ẢNH	17
2.1 Phát hiện mắt	17
2.1.1 Phát hiện mặt người	18
2.1.2. Phát hiện vị trí mắt ban đầu	19
2.2. Kiểm tra mắt sử dụng hỗ trợ máy vector	20
2.2.1. Những khái niệm cơ bản về SVM	21
2.2.1.1. Khái niệm cơ bản về SVM	21
2.2.1.2. Bài toán phân lớp	22
2.2.1.3. Phân lớp tuyến tính	22
2.2.1.4. SVM và phân cách với khoảng cách lớn nhất	23
2.2.1.5. Không gian đặc trưng	25
2.2.2 Cơ sở lý thuyết SVM	26
2.2.2.1. Bài toán phân 2 lớp với SVM	26
2.2.2.2. Bài toán phân nhiều lớp với SVM	30

2.2.2.3. Trường hợp dữ liệu không thể phân tách được.....	31
2.2.3. Huấn luyện SVM.....	33
2.2.4. Phát hiện trạng thái mất với SVM.....	34
Chương 3: KẾT QUẢ THỰC NGHIỆM	35
3.1 Bài toán	35
3.2. Các bước thực hiện	36
3.3. Giao diện chương trình và kết quả thực nghiệm	37
PHẦN KẾT LUẬN.....	41
TÀI LIỆU THAM KHẢO	43

DANH MỤC HÌNH VẼ

Hình 1.1 : Mô hình quá trình xử lý ảnh

Hình 1.2 : Mô hình việc nắn chỉnh biến dạng

Hình 1.3 Sự kết hợp mắt sơ đồ theo dõi

Hình 2.1 Ví dụ mắt người trong sang và tối

Hình 2.2 Phát hiện mắt sơ đồ khối

Hình 2.3 Hình ảnh khác biệt đôi mắt

Hình 2.4 Hình ảnh khác biệt thresholded đánh dấu ứng cử viên học sinh

Hình 2.5 Phân tách theo siêu phẳng (w,b) trong không gian 2 chiều của tập mẫu

Hình 2.6 Siêu Phẳng tối ưu

Hình 2.7 Ánh xạ Φ từ không gian dữ liệu X sang không gian đặc trưng F

Hình 2.8. Minh họa cho bài toán phân hai lớp

Hình 2.9. Minh họa bài toán phân hai lớp với phương pháp SVM

Hình 2.10. Bài toán SVM trong trường hợp dữ liệu mẫu không phân tách tuyến tính

Hình 2.11 (A) hình ảnh khác biệt thresholded chồng với ứng viên học sinh có thể. (B)

hình ảnh tối được đánh dấu với các ứng cử viên mắt có thể theo vị trí của các ứng cử viên học sinh(A)

Hình 3.1 Sự di chuyển ánh mắt của mặt người trước ảnh

Hình 3.2 Ảnh chương trình đang chạy quá trình đầu

Hình 3.3 Quá trình tiếp theo

Hình 3.4 Trạng thái mắt nhắm

Hình 3.5 Trạng thái mắt mở nhìn thẳng

Hình 3.6 Trạng thái mắt mở nhìn sang bên

Hình 3.7 Trạng thái mắt nhìn buồn ngủ

Hình 3.8 Chương trình chạy và kết quả

LỜI CẢM ƠN

Em xin được bày tỏ sự lòng biết ơn đối với thầy giáo PGS.TS. Đỗ Năng Toàn giảng viên – Viện Khoa Công nghệ thông tin. Trong suốt quá trình làm đồ án tốt nghiệp, thầy đã dành rất nhiều thời gian quý báu để hướng dẫn, chỉ bảo và định hướng cho em thực hiện đồ án của mình.

Em xin được cảm ơn các thầy, cô giáo Trường Đại học Dân lập Hải phòng đã giúp em có kiến thức lí thuyết, thực hành để em hiểu thấu đáo hơn các nội dung nghiên cứu đồ án và thực hiện đồ án này.

Em xin cảm ơn thầy, cô giáo và các bạn bè đã giúp đỡ em trong suốt quá trình học tập và đồ án tốt nghiệp.

Hải Phòng, ngày tháng năm 2013.

Sinh viên

Nguyễn Thị Hiền

MỞ ĐẦU

Trong thời đại hiện nay, sự phát triển của CNTT kéo theo sự phát triển của nhiều lĩnh vực khác, làm tăng số lượng giao dịch thông tin trên internet. Thông tin ngày một nhiều, tốc độ thay đổi cũng chóng mặt, hoạt động của các lĩnh vực cũng đặt ra xử lý một khối lượng thông tin đồ sộ. Một yêu cầu lớn đặt ra là làm thế nào để tổ chức, tìm kiếm thông tin một cách hiệu quả và phân loại thông tin là một trong những giải pháp hợp lý cho yêu cầu này. Nếu xử lý và phân loại dữ liệu thủ công là điều không tưởng với số lượng dữ liệu khổng lồ như vậy. Giải pháp được đặt ra là sử dụng máy tính để tự động phân loại các thông tin. Kỹ thuật SVM được đánh giá là công cụ mạnh và tinh vi nhất hiện nay cho những bài toán phân lớp phi tuyến. Nhiều ứng dụng đã và đang được xây dựng trên kỹ thuật SVM rất hiệu quả, trong đó có những bài toán trong lĩnh vực xử lý ảnh như phát hiện, nhận dạng mặt người.

Trong khi đó, trong lĩnh vực xử lý ảnh, các bài toán liên quan đến khuôn mặt luôn đóng vai trò quan trọng, tính thực tiễn cao.

Trong đề tài “TÌM HIỆN BÀI TOÁN PHÁT HIỆN TRẠNG THÁI MẮT CỦA MẶT NGƯỜI TRONG ẢNH” này em sẽ trình bày lý thuyết về kỹ thuật SVM, một kỹ thuật phân lớp hiện đại và hiệu quả và ứng dụng của nó trong lĩnh vực xử lý ảnh, cụ thể là bài toán phát hiện mắt của mặt người trong ảnh.

Đề tài tổ chức thành 3 chương:

Chương 1: KHÁI QUÁT VỀ XỬ LÝ ẢNH VÀ PHÁT HIỆN MẮT CỦA MẶT NGƯỜI TRONG ẢNH

Chương 2: PHÁT HIỆN MẮT CỦA MẶT NGƯỜI TRONG ẢNH

Chương 3: KẾT QUẢ THỰC NGHIỆM

Chương 1

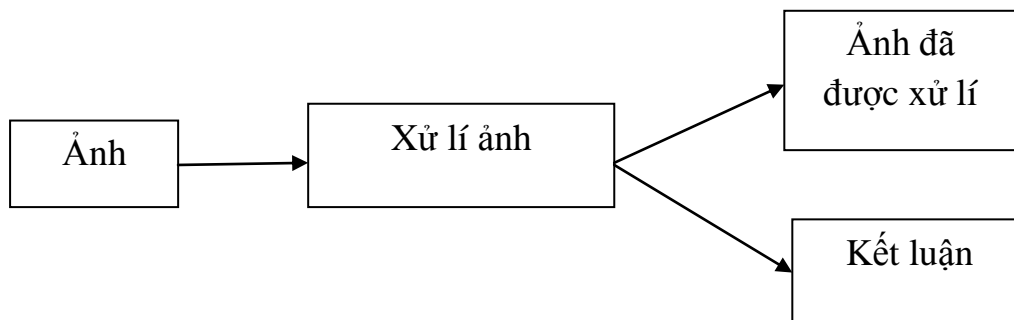
KHÁI QUÁT VỀ XỬ LÝ ẢNH VÀ PHÁT HIỆN MẶT CỦA MẶT NGƯỜI TRONG ẢNH

1.1 Khái quát về xử lý ảnh

1.1.1 Xử lý ảnh là gì?

Con người thu nhận thông tin qua các giác quan, trong đó thị giác đóng vai trò quan trọng nhất. Những năm trở lại đây với sự phát triển của phần cứng máy tính, xử lý ảnh và đồ họa đó phát triển một cách mạnh mẽ và có nhiều ứng dụng trong cuộc sống. Xử lý ảnh và đồ họa đóng một vai trò quan trọng trong tương tác người máy.

Quá trình xử lý 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 1.1 : Mô hình 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 $P(c_1, c_2, \dots, c_n)$. Do đó, ảnh trong xử lý có thể xem như n chiều.

1.1.2 Các vấn đề cơ bản trong xử lý ảnh

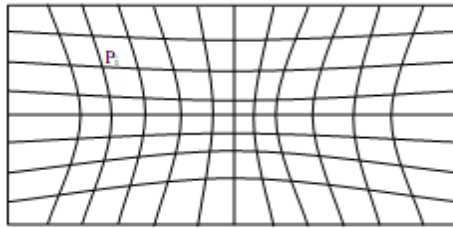
1.1.2.1 Một số khái niệm cơ bản

Ảnh và điểm ảnh : Điểm ảnh được xem như là dấu hiệu hay cường độ sáng tại một tọa độ trong không gian của đối tượng và ảnh được xem như là một tập hợp các điểm ảnh.

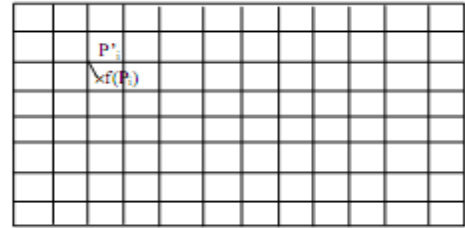
Mức xám, màu : Là số các giá trị có thể có của các điểm ảnh của ảnh.

1.1.2.2 Nắn chỉnh biến dạng

Ảnh thu nhận thường bị biến dạng do các thiết bị quang học và điện tử.



Ảnh thu nhận



Ảnh mong muốn

Hình 1.2 : Mô hình việc nắn chỉnh biến dạng

Để khắc phục người ta sử dụng các phép chiếu, các phép chiếu thường được xây dựng trên tập các điểm điều khiển.

Giả sử (P_i, P'_i) với $i=1, \dots, n$ có n các tập điều khiển

Tìm hàm $f : P_i \mapsto f(P_i)$ sao cho :

$$\sum_{i=1}^n | |f(P_i) - P'_i| |^2 \rightarrow \min$$

Giả sử ảnh bị biến đổi chỉ bao gồm : Tịnh tiến, quay, tỷ lệ, biến dạng bậc nhất tuyến tính. Khi đó hàm f có dạng :

$$f(x, y) = (a_1x + b_1y + c_1, a_2x + b_2y + c_2)$$

Ta có :

$$\phi = \sum_{i=1}^n (f(P_i) - P'_i)^2 = \sum_{i=1}^n \left[(a_1x_i + b_1y_i + c_1 - x'_i)^2 + (a_2x_i + b_2y_i + c_2 - y'_i)^2 \right]$$

Để cho $\phi \rightarrow \min$

$$\begin{cases} \frac{\partial \phi}{\partial a_1} = 0 \\ \frac{\partial \phi}{\partial b_1} = 0 \\ \frac{\partial \phi}{\partial c_1} = 0 \end{cases} \Leftrightarrow \begin{cases} \sum_{i=1}^n a_1 x_i^2 + \sum_{i=1}^n b_1 x_i y_i + \sum_{i=1}^n c_1 x_i = \sum_{i=1}^n x_i x_i' \\ \sum_{i=1}^n a_1 x_i y_i + \sum_{i=1}^n b_1 y_i^2 + \sum_{i=1}^n c_1 y_i = \sum_{i=1}^n y_i x_i' \\ \sum_{i=1}^n a_1 x_i + \sum_{i=1}^n b_1 y_i + n c_1 = \sum_{i=1}^n x_i' \end{cases}$$

Giải hệ phương trình tuyến tính tìm được a_1, b_1, c_1

Tương tự tìm được a_2, b_2, c_2

\Rightarrow Xác định được hàm f

1.1.2.3 Khử nhiễu

Có 2 loại nhiễu cơ bản trong quá trình thu nhận ảnh

- *Nhiễu hệ thống* : là nhiễu có quy luật có thể khử bằng các phép biến đổi.
- *Nhiễu ngẫu nhiên* : vết bản không rõ nguyên nhân \rightarrow khắc phục bằng các phép lọc.

1.1.2.4 Chỉnh mức xám

Nhằm khắc phục tính không đồng đều của hệ thống gây ra. Thông thường có 2 hướng tiếp cận :

- *Giảm số mức xám* : Thực hiện bằng cách nhóm các mức xám gần nhau thành một bó. Trường hợp chỉ có hai mức xám thì chính là chuyển về ảnh đen trắng. Ứng dụng : In ảnh màu ra máy in đen trắng.

- *Tăng số mức xám* : Thực hiện nội suy ra các mức xám trung gian bằng kỹ thuật nội suy. Kỹ thuật này nhằm tăng cường độ mịn cho ảnh

1.1.2.5 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. Có thể nêu ra một số đặc điểm của ảnh sau đây :

- *Đặc điểm không gian* : Phân bố mức xám, phân bố xác suất, biên độ, điểm uốn...

- *Đặc điểm biến đổi* : Các đặc điểm loại này được trích chọn bằng việc thực hiện lọc vùng (zonal filtering). Các bộ vùng được gọi là “mặt nạ đặc điểm” (feature

mask) thường là các khe hẹp với hình dạng khác nhau (chữ nhật, tam giác, cung tròn,...)

- *Đặc điểm biên và đường biên* : Đặc trưng đường biên của đối tượng rất hữu ích trong việc trích chọn các thuộc tính bất biến và được dùng khi nhận dạng đối tượng. Các đặc điểm này có thể được trích chọn nhờ toán tử gradient, toán tử la bàn, toán tử Laplace, toán tử “chéo không” (zero crossing),...

Việc 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 xuống.

1.1.2.6 Nhận dạng

Xét trên phương diện tổng quát, nhận dạng đối tượng là một công việc được thực hiện hàng ngày của những sinh vật sống và là khả năng vốn có và cần thiết của sinh vật để thích nghi với môi trường. Công việc này được thực hiện trong trong những tình huống khác nhau như là tìm kiếm nguồn thức ăn, di trú, phát hiện ra thú dữ hay là để nhận biết những người bạn v.v.. một cách rất hiệu quả. Nhận dạng đối tượng được xem như là một khái niệm nhận thức rộng nó có thể là một nhiệm vụ rất đơn giản, ví dụ như khi một vi sinh vật biến mất khỏi môi trường không đủ độ pH, hay là những nhiệm vụ phức tạp đòi hỏi khả năng suy luận, mô tả và hiểu biết nhất định, ví dụ khi một người phải tìm một cái kéo từ ngàn thứ hai đếm từ dưới lên của một cái tủ.

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. Tuy nhiên, một câu hỏi đặt ra là : mẫu (pattern) là gì? Watanabe, một trong những người đi đầu trong lĩnh vực này đã định nghĩa : “Ngược lại với hỗn loạn (chaos), mẫu là một thực thể (entity), được xác định một cách ang áng (vaguely defined) và có thể gán cho nó một tên gọi nào đó”. 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 có một mẫu nào đó, để nhận dạng hoặc phân loại mẫu đó có thể

- Hoặc phân loại có mẫu (supervised classification), chẳng hạn phân tích phân biệt (discriminant analysis), trong đó mẫu đầu vào được định danh như một thành phần của một lớp đã xác định.
- Hoặc phân loại không có mẫu (unsupervised classification hay clustering) trong đó các mẫu được gán vào các lớp khác nhau dựa trên một tiêu chuẩn đồng dạng nào đó. Các lớp này cho đến thời điểm phân loại vẫn chưa biết hay chưa được định danh.

Những ứng dụng của các hệ thống nhận dạng mẫu nói chung và hệ thống nhận dạng trong xử lý ảnh với dữ liệu hình ảnh là rất lớn và bao trùm một phạm vi rộng lớn của cuộc sống. Sau đây là một số ví dụ trong một vài hoạt động chuyên ngành :

- Nông nghiệp :
 - ✓ Phân tích cây trồng.
 - ✓ Đánh giá đất trồng.
- Thiên văn học :
 - ✓ Phân tích ảnh chụp từ kính viễn vọng.
 - ✓ Tự động hoá quang phổ học.
- Sinh học :
 - ✓ Tự động hoá tế bào học.
 - ✓ Đặc trưng của các nhiễm sắc thể.
 - ✓ Các nghiên cứu di truyền học.
- Quản lý công dân :
 - ✓ Phân tích và điều khiển luồng giao thông.
 - ✓ Định mức sự tăng trưởng của thành phố.
- Quản lý kinh tế :
 - ✓ Dự đoán thị trường chứng khoán.
 - ✓ Phân tích hiệu suất của doanh nghiệp.
- Kỹ thuật :
 - ✓ Phát hiện lỗi trong những sản phẩm được chế tạo.
 - ✓ Nhận dạng ký tự.
 - ✓ Nhận dạng tiếng nói.
 - ✓ Những hệ thống dẫn đường tự động.
 - ✓ Phân tích sự ô nhiễm.
- Địa chất :
 - ✓ Phân loại các loại đá.
 - ✓ Ước lượng những tài nguyên khai thác.
 - ✓ Phân tích nguồn tài nguyên địa chất sử dụng những hình ảnh vệ tinh.

- ✓ Phân tích địa chấn.
- Y học :
 - ✓ Phân tích điện tâm đồ.
 - ✓ Phân tích điện não đồ.
 - ✓ Phân tích những hình ảnh nội khoa.
- Quân sự :
 - ✓ Phân tích ảnh chụp không gian.
 - ✓ Phát hiện và phân loại các song ra đa và song siêu âm.
 - ✓ Tự động phát hiện mục tiêu.
- Bảo mật :
 - ✓ Phát hiện các dấu vân tay.
 - ✓ Những hệ thống giám sát và báo động.

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 bao gồm nhiều mô hình kết hợp.

Việc giải quyết bài toán nhận dạng trong những ứng dụng mới, nảy sinh trong cuộc sống không chỉ tạo ra những thách thức về thuật giải, mà còn đặt ra những yêu cầu về tốc độ tính toán. Đặc điểm chung của tất cả những ứng dụng đó là những đặc điểm đặc trưng cần thiết thường là nhiều, không thể do chuyên gia đề xuất, mà phải được trích chọn dựa trên các thủ tục phân tích dữ liệu.

1.1.2.7 Nén ảnh

Nhằm giảm thiểu không gian lưu trữ. Thường được tiến hành theo cả hai cách khuynh hướng là nén có bảo toàn và không bảo toàn thông tin. Nén không bảo toàn thì thường có khả năng nén cao hơn nhưng khả năng phục hồi thì kém hơn. Trên cơ sở hai khuynh hướng, có 4 cách tiếp cận cơ bản trong nén ảnh :

- Nén ảnh thống kê : Kỹ thuật nén này dựa vào việc thống kê tần suất xuất hiện của giá trị các điểm ảnh, trên cơ sở đó mà có chiến lược mã hoá thích hợp. Một ví dụ điển hình cho kỹ thuật mã hoá này là *.TIF

- Nén ảnh không gian : Kỹ thuật này dựa vào vị trí không gian của các điểm ảnh để tiến hành mã hoá. Kỹ thuật lợi dụng sự giống nhau của các điểm ảnh trong các vùng gần nhau. Ví dụ cho kỹ thuật này là mã nén *.PCX

- Nén ảnh sử dụng phép biến đổi : Đây là kỹ thuật tiếp cận theo hướng nén không bảo toàn và do vậy, kỹ thuật thường nén hiệu quả hơn. *.JPG chính là tiếp cận theo kỹ thuật nén này.

- Nén ảnh Fractal : Sử dụng tính chất Fractal của các đối tượng ảnh, thể hiện sự lặp lại của các chi tiết. Kỹ thuật nén sẽ tính toán để chỉ cần lưu trữ phần gốc ảnh và quy luật sinh ra ảnh theo nguyên lý Fractal.

1.2 Phát hiện trạng thái mắt của mặt người trong ảnh

1.2.1 . Giới thiệu

Một vấn đề quan trọng trong hệ thống nhận dạng khuôn mặt là mặt gắn kết triền. Mặt liên kết liên quan đến không gian rộng và luân phiên một hình ảnh khuôn mặt để phù hợp với hình ảnh khuôn mặt trong cơ sở dữ liệu. Nó đã chỉ ra rằng sự liên kết khuôn mặt có tác động lớn về tính chính xác công nhận [17, 15].

Hiện nay, liên kết mặt thường được thực hiện với việc sử dụng các vị trí mắt. Đối với hầu hết phải đối mặt với các phương pháp nhận dạng, vị trí mắt được tự nhất định. Nhưng đối với một ứng dụng thế giới thực của nhận dạng khuôn mặt, manu- đồng minh phát hiện các vị trí mắt rõ ràng là không thực tế. Một do đó thuật toán tự động phát hiện mắt là cần thiết cho một hệ thống nhận diện khuôn mặt hoàn toàn tự động.

Trong bài báo này, chúng tôi đầu tiên đề xuất một thời gian thực mới tự động phương pháp phát hiện mắt. Phương pháp phát hiện mắt của chúng tôi là sau đó xác nhận sử dụng cơ sở dữ liệu FRGC [16]. Phần còn lại của bài viết này được tổ chức như sau: tác động của vị trí mắt trên khuôn mặt nhận được thảo luận trong phần 2. Các công việc liên quan trên phát hiện mắt tự động được xem xét trong phần 3. Chúng tôi ủng hộ đặt ra một thuật toán nội địa hóa mắt chính xác tại Mục 4. Trong Phần 5, chúng tôi trình bày các kết quả thí nghiệm chứng thực của chúng tôi nội địa hóa mắt cho nhận dạng khuôn mặt trên các cơ sở dữ liệu FRGC. Bài viết kết luận trong phần 6.

1.2.2. Phát hiện 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.

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

- 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 khác nhau trên khuôn mặt, như: vui, buồn, ngạc nhiên,.....
- Mặt người bị che khuất bởi các đối tượng khác có trong ảnh.
- Sự biểu cảm của khuôn mặt: sự biểu cảm có thể làm thay đổi đáng kể các đặc trưng và thông số khuôn mặt, ví dụ như khuôn mặt của cùng một người sẽ rất khác khi người ấy cười, tức giận hay sợ hãi.....

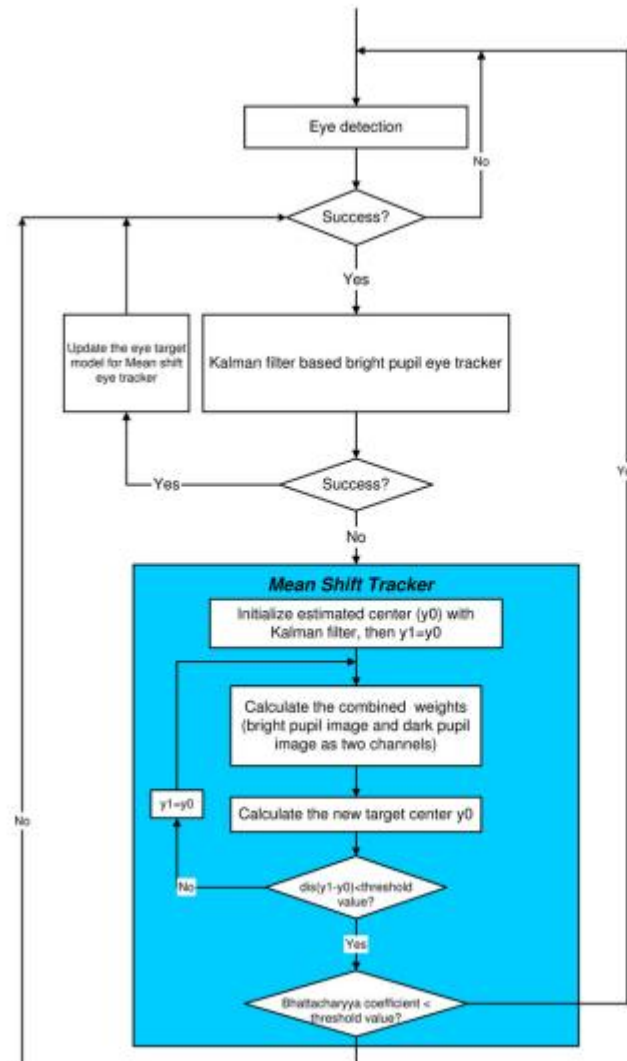
Xác định mặt người thường là một phần của hệ thống. Nó thường được dùng trong giám sát video, giao tiếp người máy và quản lý cơ sở dữ liệu ảnh.... Các ứng dụng cơ bản của xác định mặt người có thể kể đến.

1.2.3. Phát hiện mắt người trong ảnh

Với đôi mắt phát hiện trong khung ban đầu, đôi mắt trong khung tiếp theo có thể được theo dõi từ frame này sang frame. Theo dõi mắt có thể được thực hiện bằng cách thực hiện học sinh phát hiện trong mỗi khung hình. Phương pháp bạo lực này, tuy nhiên, sẽ làm chậm đáng kể tốc độ của việc theo dõi học sinh, làm cho thời gian thực theo dõi học sinh không thể vì nó cần phải tìm kiếm toàn bộ hình ảnh cho mỗi khung hình. Điều này có thể được thực hiện hiệu quả hơn bằng cách sử dụng các chương trình của báo và phát hiện. Lọc Kalman [42] cung cấp một cơ chế để thực hiện điều này. Các Kalman học sinh theo dõi, tuy nhiên, có thể thất bại nếu học sinh không đủ sáng theo các điều kiện đã đề cập.

Ngoài ra, phong trào đầu nhanh chóng cũng có thể gây ra các tracker để mắt đôi mắt. Vấn đề này được giải quyết bằng cách gia tăng sự Kalman theo dõi với sự thay đổi tracker. Fig trung bình. 13 tóm tắt chương trình theo dõi mắt của chúng tôi. Cụ thể, sau khi định vị mắt trong các khung ban đầu, lọc Kalman được kích hoạt để theo dõi học sinh sáng. Nếu nó không thành công trong một khung do phân xuất hiện của học sinh sáng, theo dõi mắt dựa trên sự thay đổi trung bình sẽ lên ngôi.

Theo dõi mắt của chúng tôi sẽ trở lại để theo dõi học sinh sáng ngay sau khi học sinh sáng xuất hiện trở lại kể từ khi nó là mạnh mẽ hơn nhiều và theo dõi đáng tin cậy. Phát hiện học sinh sẽ được kích hoạt nếu theo dõi sự thay đổi có nghĩa là thất bại. Hai máy theo dõi mắt giai đoạn làm việc cùng nhau và họ bổ sung cho nhau. Sự vững mạnh của tracker mắt được cải thiện đáng kể.



Hình 1.3 Sự kết hợp mắt sơ đồ theo dõi

1.2.4. Phát hiện trạng thái mặt người theo một số phương pháp

Sirohey đề xuất phương pháp phát hiện và định vị khuôn mặt từ ảnh có nền phức tạp [4]. Ông đã tìm ra các cạnh (được phát hiện theo phương pháp Candy), sau đó loại bỏ bớt và nhóm các cạnh lại sao cho chỉ còn lại một biên bao quanh khuôn mặt. Sau đó hình thành một elip vừa với biên vừa rồi phân biệt vùng đầu và nền. Giải thuật này đạt độ chính xác 80% trên bộ 48 ảnh với nền phức tạp. Thay vì dùng cạnh, Chetverikov và Lerch dùng các đốm và vạch sọc để phát hiện khuôn mặt [5]. Hai ông dùng 2 đốm tối và 3 đốm sáng để tả 2 mắt, 2 gò má và mũi, dùng các vạch sọc để thể hiện đường nét của khuôn mặt, lông mày và đôi môi. Sử dụng hai hình tam giác để mô tả quan hệ giữa các đốm. Một biến đổi Laplace của ảnh được sinh ra để xác định các đốm, sau đó quét hết ảnh để tìm chính xác sự tồn tại của các hình tam giác xem như ứng với các ứng viên. Khuôn mặt được phát hiện nếu các vạch sọc coi là bao quanh ứng viên.

Graf đã dùng một phương pháp để định vị các đặc trưng khuôn mặt và khuôn mặt đối với ảnh đa mức xám [6]. Tư tưởng như sau: đầu tiên cho ảnh qua một bộ lọc để làm nổi các biên, tiếp theo dùng các phép hình thái học để làm nổi bật các vùng có cường độ cao và hình dạng rõ ràng. Histogram của ảnh bây giờ sẽ chứa một đỉnh khá nổi, dựa vào vị trí đó, xác định các giá trị ngưỡng phù hợp để tạo ra 2 ảnh nhị phân tương ứng. Các phần có liên hệ giữa cả hai ảnh nhị phân này sẽ xác định vùng ứng viên khuôn mặt. Kết hợp các vùng này sẽ được xem xét với các bộ phân loại, từ đó xem xét ứng viên có đúng là khuôn mặt hay không. Tuy nhiên phương pháp này gặp vấn đề với việc dùng các phép toán hình thái học như đã nói trên đây, thêm nữa là việc kết hợp các đặc trưng để định vị khuôn mặt.

Leung đã phát triển một phương pháp để định vị khuôn mặt trong ảnh nền phức tạp dựa vào việc xác định các đặc trưng cục bộ và bộ khớp đồ thị ngẫu nhiên. Mục đích là đưa ra quy tắc định vị khuôn mặt bằng việc tìm ra cách sắp xếp các đặc trưng sao cho gần nhất với mẫu khuôn mặt [7]. Dùng 5 đặc trưng : 2 mắt, 2 lỗ mũi, đoạn giữa môi và mũi được dùng để thể hiện một khuôn mặt chuẩn. Với mỗi cặp đặc trưng này, tính khoảng cách liên hệ giữa chúng, tập các khoảng cách đó được mô hình bằng mô hình Gauss. Định nghĩa một mẫu khuôn mặt bởi giá trị trung bình tập kết quả của bộ lọc đạo hàm đa hướng, đa tỉ lệ (tại các pixel trong vùng đặc trưng mặt). Với 1 ảnh cần kiểm tra, mỗi đặc trưng mặt được chỉ ra bằng cách khớp kết quả qua bộ lọc tại mỗi pixel với véc tơ đáp ứng mẫu. Chọn hai đặc trưng với đáp ứng tốt nhất để làm cơ sở cho việc tìm các đặc trưng khác của khuôn mặt. Vì rằng các đặc trưng không thể xuất hiện tùy tiện nên cần dùng mô hình xác suất để định vị chúng qua khoảng cách. Thêm nữa việc tính toán hiệp phương sai cùng giúp việc định vị đặc trưng được tốt hơn. Các hình sao được hình thành từ các đặc trưng ứng viên trong vùng thích hợp. Việc tìm kiếm hình sao tốt nhất được xem như việc khớp đồ thị mà các nút tương ứng với các đặc trưng khuôn mặt, còn các cạnh ứng với khoảng cách giữa các đặc trưng. Hạng của các hình sao được tính dựa trên hàm tỉ lệ xác suất mà hình sao đó ứng với khuôn mặt thật và không phải khuôn mặt.

Han đã phát triển một kĩ thuật hình thái học để trích những vùng mà ông gọi là giống hay tương tự với mắt phục vụ việc phát hiện khuôn mặt [8]. Tư tưởng trên cơ sở ý tưởng sau: mắt và lông mày là đặc trưng ổn định nhất của con người. Ông định nghĩa vùng tương tự mắt là các cạnh viền quanh mắt. Đầu tiên dùng các phép toán hình thái học như đóng, bớt sai khác, chia ngưỡng để trích các pixel có mức xám có thay đổi đáng kể. Những pixel này trở thành các pixel trong vùng họ định nghĩa như trên. Tiếp theo là giai đoạn gán chúng thành các vùng tương tự mắt. Trên cơ sở sử dụng các vùng này và kết hợp các đặc trưng hình học của mắt, mũi, lông mày và miệng. Cuối cùng dùng một mạng nơron để xem xét các ứng viên để đưa ra kết quả.

Chương 2

PHÁT HIỆN MẮT CỦA MẶT NGƯỜI TRONG ẢNH

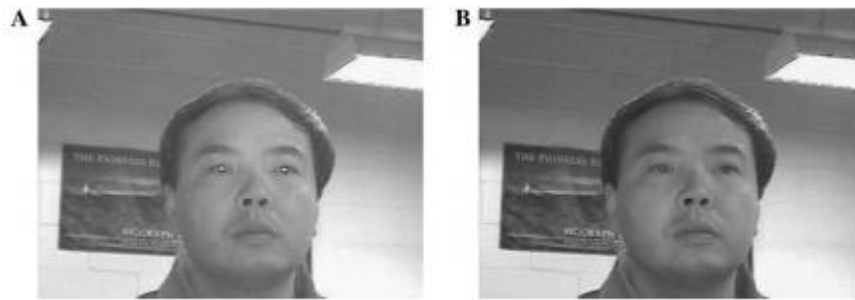
2.1 Phát hiện mắt

Để tạo điều kiện xử lý hình ảnh tiếp theo, khuôn mặt của người được chiếu sáng sử dụng đèn chiếu sáng cận hồng ngoại. Việc sử dụng đèn hồng ngoại phục vụ ba mục đích:

- Đầu tiên nó giảm thiểu tác động của điều kiện ánh sáng môi trường xung quanh khác nhau, do đó đảm bảo chất lượng của tuổi theo các điều kiện thực tế bao gồm chiếu sáng ngày và đêm.
- Thứ hai: Nó cho phép để sản xuất sáng / tối có hiệu lực, mà là nền tảng để phát hiện mắt được đề xuất và thuật toán theo dõi.
- Thứ ba: Khi gần hồng ngoại thì gần như không thể cho người sử dụng, điều này sẽ giảm thiểu bất kỳ sự can thiệp với người sử dụng? Theo bằng sáng chế ban đầu từ Hutchinson [39], một người bình thường có thể thu được nếu mắt được chiếu sáng với đèn hồng ngoại chiếu sáng ánh sáng gần dọc theo trục quang của máy ảnh ở bước sóng nhất định.

Ở gần hồng ngoại bước sóng, người thường phản ánh gần như tất cả ánh sáng hồng ngoại mà họ nhận được trên con đường trở lại máy ảnh, kết quả ảnh hưởng tới ánh sáng tới mắt người, rất nhiều tương tự như hiệu ứng mắt đỏ trong nhiếp ảnh. Nếu chiếu sáng ngoài trục quang của máy ảnh, nhiều người xuất hiện trong bóng tối kể từ khi ánh sáng phản xạ sẽ không nhập vào ống kính máy ảnh. Điều này tạo ra cái gọi là hiệu ứng bóng tối. Một ví dụ của nhiều người trong sáng / tối được đưa ra in Fig. 4. Thông tin chi tiết về việc xây dựng các đèn chiếu sáng hồng ngoại và cấu hình của nó có thể được tìm thấy trong [40]. Với những hình ảnh mắt chiếu sáng hồng ngoại, phát hiện mắt được thực hiện thông qua phát hiện nhiều người. Phát hiện nhiều người được thực hiện dựa trên cả hai cường độ của con người (độ sáng và tối) và sự xuất hiện của đôi mắt bằng cách sử dụng hỗ trợ máy vector.

Cụ thể, phát hiện con người bắt đầu với tiền xử lý để can thiệp chiếu sáng loại bỏ bề ngoài, tiếp theo là tìm kiếm trên toàn bộ hình ảnh cho người về cường độ và sự xuất hiện mắt. Do đó, nhiều người có thể được huỷ bỏ được bảo vệ nếu có tồn tại nhiều hơn một người. Việc sử dụng các máy hỗ trợ vector (SVM) tránh sai xác định một khu vực sáng như một pupil. Fig. 5 gives một tổng quan về các module phát hiện mắt.



Hình 2.1 Ví dụ mắt người trong sang và tối

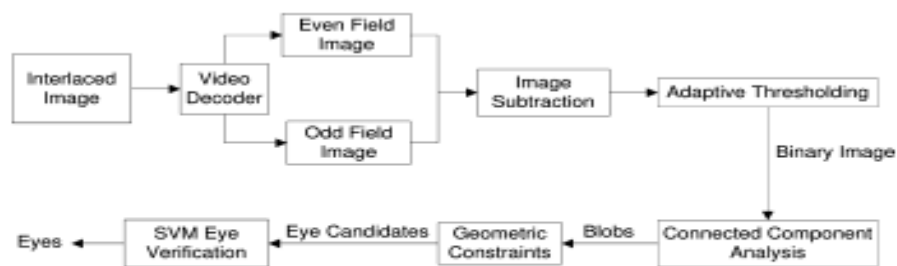


Fig. 5. Eye detection block diagram.

Hình 2.2 Phát hiện mắt sơ đồ khối

2.1.1 Phát hiện mặt người

Có nhiều nghiên cứu tìm phương pháp xác định khuôn mặt người, từ ảnh xám đến ảnh màu. Dựa vào tính chất của các phương pháp xác định khuôn mặt trên ảnh, chúng ta có thể phân chia các phương pháp này thành bốn hướng tiếp cận chính:

- Hướng tiếp cận dựa trên tri thức: Dựa vào các thuật toán, mã hóa các đặc trưng và quan hệ giữa các đặc trưng của khuôn mặt thành các luật. Đây là hướng tiếp cận theo kiểu top-down.
- Hướng tiếp cận dựa trên đặc trưng không thay đổi: Xây dựng các thuật toán để tìm các đặc trưng của khuôn mặt mà các đặc trưng này không thay đổi khi tư thế khuôn mặt hay vị trí đặt camera thay đổi.
- Hướng tiếp cận dựa trên so sánh khớp mẫu: Dùng các mẫu chuẩn của khuôn mặt (các mẫu này đã được chọn và lưu trữ) để mô tả các khuôn mặt hay các đặc trưng của khuôn mặt (các mẫu này được chọn tách biệt theo tiêu chuẩn đã được các tác giả đề ra để so sánh). Phương pháp này có thể dùng để xác định vị trí hay dò tìm khuôn mặt trên ảnh.
- Hướng tiếp cận dựa trên diện mạo: Trái ngược với hướng tiếp cận dựa trên khuôn mẫu, các mô hình (hay các mẫu) sẽ được học từ một tập ảnh huấn luyện mà thể hiện tính chất tiêu biểu của sự xuất hiện của mặt người trong ảnh. Sau đó hệ thống (mô hình) sẽ xác định mặt người. Hướng tiếp cận này còn được biết là hướng tiếp cận theo các phương pháp học máy.

Hướng tiếp cận trong luận văn

Luận văn sẽ tiếp cận theo hướng sử dụng một phương pháp trích chọn đặc trưng ảnh kết hợp với SVM (Support vector machines).

-Bước 1, chiếu ảnh khuôn mặt từ không gian ảnh thô sang không gian các không gian khuôn mặt (Mỗi lớp khuôn mặt được nhận dạng sẽ được mô hình hóa bằng một không gian khuôn mặt) dùng kỹ thuật trích chọn đặc trưng ảnh.

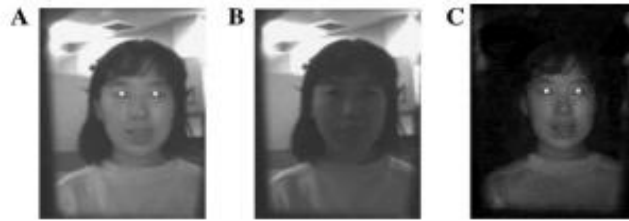
-Bước 2, sử dụng SVM để tạo bộ phân loại tuyến tính có khả năng quyết định phân lớp.

2.1.2. Phát hiện vị trí mắt ban đầu

Các thuật toán phát hiện bắt đầu với một tiền xử lý để giảm thiểu sự can thiệp từ các nguồn chiếu sáng khác với các đèn chiếu sáng hồng ngoại. Điều này bao gồm ánh sáng mặt trời và AMBI-ent giao thoa ánh sáng. Một phương pháp khác biệt được sử dụng để loại bỏ các nền lệch giữa bằng cách trừ đi các hình ảnh mắt tối (trường lẽ) từ hình ảnh mắt sáng (ngay cả trường), tạo ra một hình ảnh khác biệt, với hầu hết các nền và bên ngoài như Hình. 4. Sáng (A) và tối (B) học sinh hình ảnh. Và. 5.

Mắt phát hiện sơ đồ khối để theo dõi mắt theo thời gian thực, phép trừ hình ảnh phải được thực hiện có hiệu quả trong thời gian thực. Để đạt được điều này, chúng tôi phát triển mạch để đồng bộ hóa các vòng ngoài của đèn LED và vòng trong của đèn LED với các lĩnh vực và thậm chí kỳ lạ của hình ảnh xen kẽ, tương ứng để họ có thể được bật và tắt luân phiên. Khi lĩnh vực thậm chí còn đang được quét, vòng trong của đèn LED là trên và vòng ngoài của đèn LED là tắt và ngược lại khi thậm chí nộp được quét. Những hình ảnh đầu vào xen kẽ là sau đó để xen kẽ thông qua một đoạn video de-coder, sản xuất những hình ảnh ngay cả lĩnh vực và lẽ như inFigs hiển thị. 6A và B. Thông tin thêm về hình ảnh của chúng tôi trừ mạch có thể được tìm thấy trong [40].

Những hình ảnh khác biệt là sau đó thresholded tự động dựa trên của nó ảnh nhị phân tạo ra một hình ảnh nhị phân. Sau đó phân tích thành phần kết nối được áp dụng cho các hình ảnh nhị phân để xác định các file nhị nguyên phân. Nhiệm vụ của chúng tôi là sau đó để tìm ra những đốm màu thực sự là nhiều người blob thực. Ban đầu, chúng tôi đánh dấu tất cả các đốm màu như các ứng cử viên tiềm năng cho 1 người như inFig hiển thị. 7. Và. 6. Nền chiếu sáng loại bỏ sự can thiệp (A) các lĩnh vực hình ảnh thậm chí còn thu được với cả môi trường xung quanh và ánh sáng hồng ngoại, (B) các lĩnh vực hình ảnh lẽ thu được với ánh sáng môi trường xung quanh chỉ; (C) các hình ảnh kết quả của trừ (B) từ (A)



Hình 2.3 Hình ảnh khác biệt đôi mắt

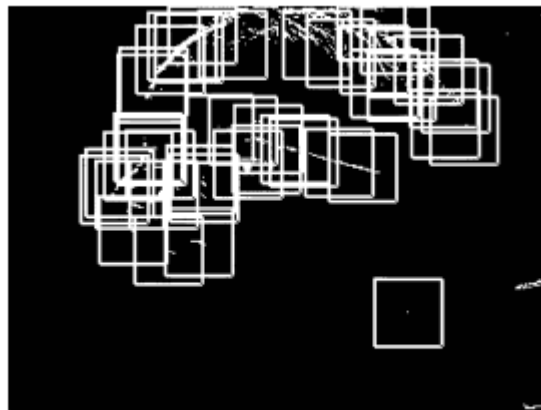


Fig. 7. The thresholded difference image marked with pupil candidates.

Hình 2.4 Hình ảnh khác biệt thresholded đánh dấu ứng cử viên học sinh

2.2. Kiểm tra mắt sử dụng hỗ trợ máy vector

Như thể hiện inFig. 7, thường có nhiều ứng cử viên tiềm năng của người. Thông thường, học sinh được phát hiện trong các file nhị nguyên phân. Tuy nhiên, nó thường là không thể để cô lập các học sinh blob chỉ bằng cách chọn các giá trị ngưỡng đúng, vì học sinh thường nhỏ và không đủ sáng so với các đốm màu tiếng ồn khác. Do đó, chúng tôi sẽ phải sử dụng các thông tin khác hơn cường độ để xác định chính xác họ. Một trong những cách ban đầu để phân biệt các đốm màu học sinh từ các đốm màu tiếng ồn khác là dựa trên hình dạng hình học của họ.

Thông thường, học sinh là một đốm hình elip giống như chúng ta có thể sử dụng một phụ kiện hình elip phương pháp [41] để trích xuất các hình dạng của mỗi blob và sử dụng các hình dạng và kích thước để loại bỏ một số đốm màu xem xét thêm. Tuy nhiên cần lưu ý do thay đổi quy mô (khoảng cách đến máy ảnh) và thay đổi kích thước học sinh cá nhân, kích thước không phải là một tiêu chuẩn đáng tin cậy. Nó chỉ được sử dụng để loại bỏ các đốm màu rất lớn hoặc rất nhỏ. Hình dạng tiêu chuẩn, mặt khác là quy mô bất biến. Tuy nhiên, hình dáng một mình là không đủ đưa ra các đốm màu không học sinh khác với hình dạng và kích thước tương tự như inFig. 8, nơi mà chúng ta có thể thấy rằng vẫn còn có một số các đốm màu

không học sinh lại bởi vì họ là như vậy tương tự như trong hình dạng và kích thước mà chúng ta không thể phân biệt thực sự đốm màu học sinh từ họ. Vì vậy, chúng ta phải sử dụng các tính năng khác. Chúng tôi quan sát thấy rằng mắt lại Gion xung quanh học sinh có phân bố cường độ độ chói.

Chúng xuất hiện khác nhau từ các bộ phận khác của khuôn mặt trong hình ảnh học sinh tối như inFig hiển thị. 4B. Xuất hiện hiem mắt do đó có thể được sử dụng để tách nó từ không mắt. Nhiều bản đồ vị trí của các file nhị nguyên phân còn lại cho hình ảnh học sinh tối và sau đó áp dụng hỗ trợ máy vector (SVM) phân loại [36,37] để tự động xác định các đốm màu bin phân tương ứng với đôi mắt như thảo luận dưới đây

2.2.1. Những khái niệm cơ bản về SVM

2.2.1.1. Khái niệm cơ bản về SVM

SVM là phương pháp học có giám sát do Vladimir N. Vapnik đề xuất vào năm 1995, và ngày càng được sử dụng phổ biến trong nhiều lĩnh vực, đặc biệt là lĩnh vực phân loại mẫu và nhận dạng mẫu.

Đề phân lớp (Classification) và dự đoán (prediction) là hai bài toán cơ bản và có rất nhiều ứng dụng trong tất cả các lĩnh vực. Có nhiều phương pháp đã được nghiên cứu và ứng dụng cho các bài toán dạng này như: mạng Neron nhân tạo, phương pháp học thống kê,... Trong chương này, chúng ta sẽ đi sâu nghiên cứu phương pháp Support Vector Machines, một phương pháp rất hiệu quả hiện nay.

SVM được coi là công cụ mạnh cho những bài toán phân lớp phi tuyến. Phương pháp này thực hiện phân lớp dựa trên nguyên lý cực tiểu hóa rủi ro có cấu trúc SRM (Structural Risk Minimization), được xem là một trong các phương pháp phân lớp giám sát không tham số tinh vi nhất cho đến nay. Các hàm công cụ đa dạng của SVM cho phép tạo không gian chuyển đổi để xây dựng mặt phẳng phân lớp.

SVM dạng chuẩn lấy một tập hợp các dữ liệu đầu vào và dự báo mỗi dữ liệu đầu vào ứng với một lớp (class) trong số hai lớp mà dữ liệu ấy có khả năng rơi vào. Điều này làm cho dạng chuẩn SVM trở nên không có tính xác suất mà là một công cụ binary tuyến tính.

Bài toán phân lớp, sử dụng SVM nhằm mục đích tìm một siêu phẳng có biên cực đại giữa lớp mẫu âm và mẫu dương, đồng thời cực tiểu hóa các mẫu không phân chia được trong tập huấn luyện. SVM dựa trên cơ sở toán học vững chắc. Tuy nhiên việc huấn luyện mẫu sử dụng SVM đòi hỏi phải giải bài toán tối ưu nhiều biến. Ban đầu, SVM được phát triển để giải các bài toán phân lớp, về sau do tính ưu việt, nó còn được ứng dụng rộng rãi để giải các bài toán hồi quy

2.2.1.2. Bài toán phân lớp

Phân lớp là quá trình nhóm các đối tượng “giống” nhau vào “một lớp” dựa trên các đặc trưng dữ liệu của chúng. Tuy nhiên, phân lớp là một hoạt động tiềm ẩn trong tư duy con người khi nhận dạng thế giới thực, đóng vai trò quan trọng làm cơ sở đưa ra các dự báo, các quyết định. Phân lớp và cách mô tả các lớp giúp cho tri thức được định dạng và lưu trữ trong đó.

Khi nghiên cứu một đối tượng, hiện tượng, chúng ta chỉ có thể dựa vào một số hữu hạn các đặc trưng của chúng. Nói cách khác, ta chỉ xem xét biểu diễn của đối tượng, hiện tượng trong một không gian hữu hạn chiều, mỗi chiều ứng với một đặc trưng được lựa chọn. Khi đó, phân lớp dữ liệu trở thành phân hoạch dữ liệu thành các tập con theo một chuẩn nhận dạng được.

Một cách hình thức, xem $X = \{(x_1, x_2, \dots, x_n)\}$ là không gian biểu diễn n-chiều các đặc trưng của đối tượng, hiện tượng. Khi đó, mỗi đối tượng của X sẽ thuộc về một lớp C_k nào đó và các lớp C_k tạo thành một phân hoạch trên X, gọi là $C(X)$

$$C(X) = \{C_k | k = \overline{1, K} \text{ và } C_k \cap C_j = \emptyset, \forall k \neq j\}$$

Vậy, việc phân lớp các đối tượng là xác định một hàm $f: X \rightarrow C(X)$, và bài toán phân lớp có thể được mô tả một cách hình thức như sau:

Cho tập mẫu $\Omega = \{(X_i, C_i) | X_i \in R^n; C_i \in R^m; i = \overline{1, N}\}$

Tìm ánh xạ $f: R^n \rightarrow R^m$, sao cho $\|f(X_i) - C_i\| \cong 0, \forall (X_i, C_i) \in \Omega, \forall i = \overline{1, N}$

Trong đó:

N: số mẫu;

X_i : mẫu dữ liệu thứ i

C_i : lớp của mẫu dữ liệu thứ i

Như vậy việc giải bài toán phân lớp là quá trình xây dựng một mô hình (các hàm, các luật...) để quyết định mỗi đối tượng, vật thể thuộc vào lớp nào dựa trên các đặc trưng dữ liệu của chúng. Quá trình phân lớp sẽ trở nên khó khăn hơn khi đối tượng có nhiều đặc trưng dữ liệu, và quá trình phân lớp cũng cần xác định đặc trưng nào cần thiết cho việc phân lớp, đặc trưng nào dư thừa

2.2.1.3. Phân lớp tuyến tính

Việc phân lớp nhị phân được thực hiện bằng cách sử dụng hàm giá trị thực $f: \subseteq R^n \rightarrow R$ là hàm tuyến tính, tương ứng với đầu ra $y \in \{-1, 1\}$, được phát biểu như sau:

Đầu vào $x=(x_1,x_2,\dots x_n)$ sẽ được gán vào lớp có nhãn 1 nếu $f(x) \geq 0$, trường hợp ngược lại sẽ được gán vào lớp có nhãn -1. Trường hợp $f(x)$ là hàm tuyến tính của $x \in X$ ta có thể viết như sau:

$$f(x) = \langle w, x \rangle + b = \sum_{i=1}^n w_i x_i + b$$

Trong đó $\langle . \rangle$ biểu thị tích vô hướng.

Về mặt hình học, các phần tử của không gian đầu vào X sẽ được rơi vào một trong hai phần được phân tách bởi siêu phẳng xác định bằng biểu thức

$$\langle w, x \rangle + b = 0$$

Trong đó: w là vector pháp tuyến của siêu phẳng, giá trị của ngưỡng b thay đổi có thể tạo ra các siêu phẳng song song với nhau.

Với mỗi mẫu dữ liệu x chưa xác định sẽ được phân chia thành:

$$x \in \begin{cases} \text{Lớp I, nếu } f(x) > 0 \\ \text{Lớp II, trong các trường hợp còn lại} \end{cases}$$

2.2.1.4. SVM và phân cách với khoảng cách lớn nhất

Support Vector Machines (SVM) là kỹ thuật mới đối với việc phân lớp dữ liệu, là phương pháp học sử dụng không gian giả thuyết các hàm tuyến tính trên không gian đặc trưng nhiều chiều, dựa trên lý thuyết tối ưu và lý thuyết thống kê.

Trong kỹ thuật SVM không gian dữ liệu nhập ban đầu sẽ được ánh xạ vào không gian đặc trưng và trong không gian đặc trưng này mặt siêu phẳng phân chia tối ưu sẽ được xác định.

Ta có tập S gồm l các mẫu học:

$$S = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots (x_l, y_l)\} \subseteq (X \times Y)^l,$$

Với một vecto đầu vào n chiều $x_i \in R^n$ thuộc lớp I hoặc lớp II (tương ứng nhãn $y_i=1$ đối với lớp I và $y_i=-1$ đối với lớp II). Một tập mẫu học được gọi là tầm thường nếu tất cả các nhãn là bằng nhau.

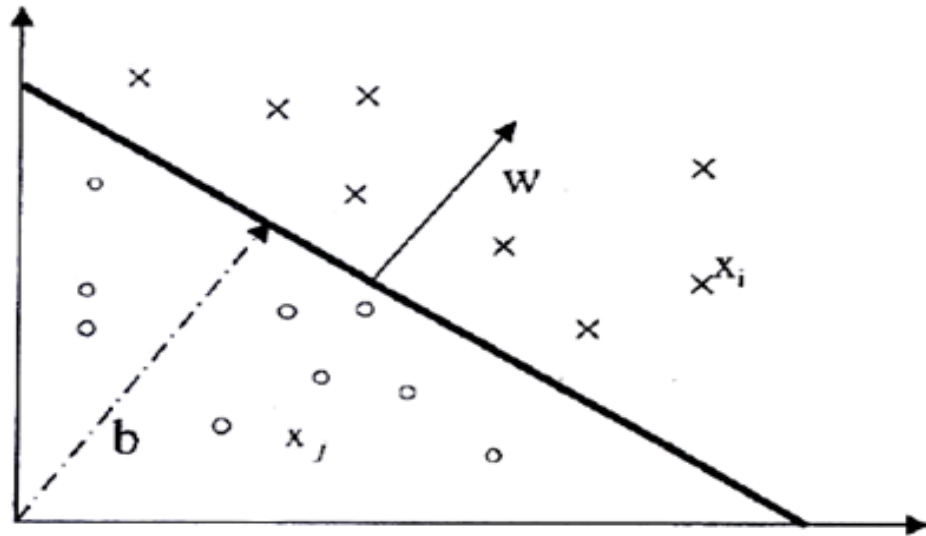
Đối với các dữ liệu phân chia tuyến tính, chúng ta có thể xác định được siêu phẳng $f(x)$ mà nó có thể chia tập dữ liệu. Khi đó, với một siêu phẳng nhận được ta có: $f(x) \geq 0$ nếu đầu vào x thuộc lớp dương, và $f(x) < 0$ nếu x thuộc lớp âm.

$$f(x) = wx + b = \sum_{j=1}^n w_j x_j + b \tag{2.1}$$

$$y_i f(x_i) = y_i (w \cdot x_i + b) \geq 0, \quad i = 1, \dots, l \tag{2.2}$$

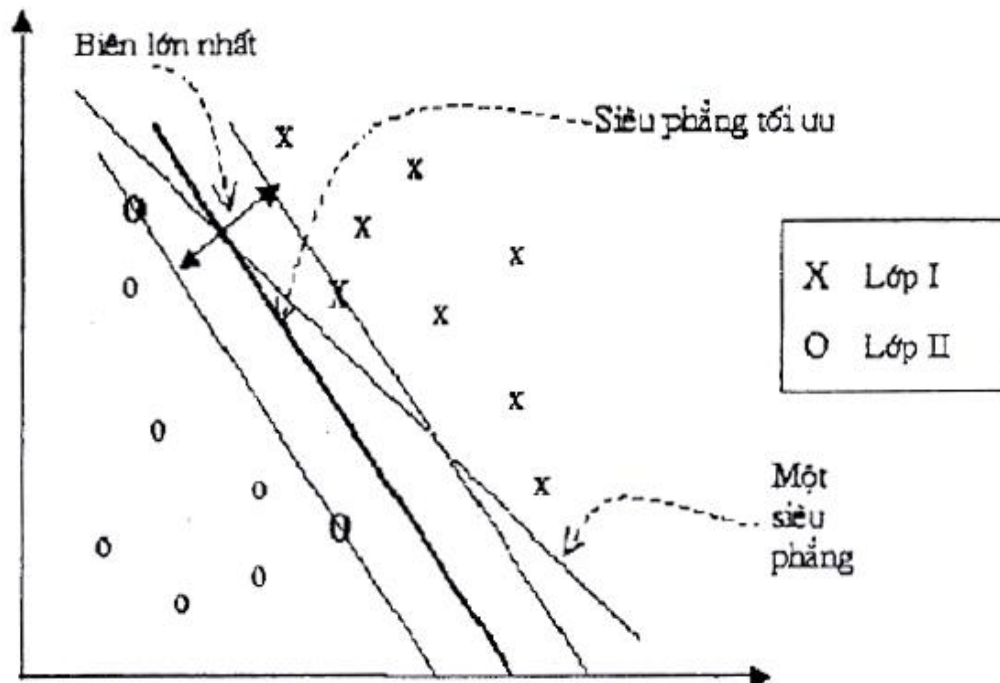
Trong đó w là vecto pháp tuyến n chiều và b là giá trị ngưỡng.

Vecto pháp tuyến w xác định chiều của siêu phẳng $f(x)$, còn giá trị ngưỡng b xác định khoảng cách giữa siêu phẳng và gốc.



Hình 2.5 Phân tách theo siêu phẳng (w,b) trong không gian 2 chiều của tập mẫu

Siêu phẳng có khoảng cách với dữ liệu gần nhất là lớn nhất (tức có biên lớn nhất) được gọi là siêu phẳng tối ưu. Một ví dụ về siêu phẳng tối ưu phân chia loại dữ liệu được mô tả trong hình 2.2 sau đây:



Hình 2.6 Siêu Phẳng tối ưu

Mục đích đặt ra ở đây là tìm được một ngưỡng (w,b) phân chia tập mẫu vào các lớp có nhãn 1 (lớp I) và -1 (lớp II) nêu ở trên với khoảng cách là lớn nhất.

2.2.1.5. Không gian đặc trưng.

SVM là một công cụ phân chia phi tuyến dựa trên hàm hạt nhân (kernel), nó chuyển đổi dữ liệu được phân chia x sang một không gian mới, mà tại đó các mẫu dữ liệu này có thể phân chia một cách tuyến tính. Không gian này gọi là không gian đặc trưng. Bằng việc sử dụng hàm vecto phi tuyến $\Phi(x) = (\Phi_1(x), \dots, \Phi_l(x))$ đã chuyển vecto đầu vào n chiều x sang một không gian đặc trưng l chiều, hàm tuyến tính ở cả hai miền sẽ là:

$$f(x) = \sum_{i=1}^l \alpha_i y_i \Phi(x_i) \cdot \Phi(x) \tag{2.3}$$

Đối với SVM, việc tính hàm chuyển đổi Φ là không cần thiết nhưng các lớp của siêu diện có thể tính bằng tích trong của các mẫu dữ liệu gốc. Nếu có thể xác định được cách tính tích trong của không gian đặc trưng trong không gian dữ liệu gốc thì đó gọi là một hàm Kernel, $K(x, z) = \Phi(x) \cdot \Phi(z)$. Do vậy việc học trong không gian đặc trưng không yêu cầu phải xác định Φ vì tất cả các mẫu gốc đã được đơn ánh với ma trận Gram $G = \{(x_i \cdot x_j)\}_{i,j=1}^M$. Sử dụng hàm Kernel, hàm quyết định trở thành:

$$f(x) = \text{sign}\left(\sum_{i \in \text{support vector}} \alpha_i \times y_i K(x_i, x)\right)$$

Và các mẫu dữ liệu không xác định được chia thành:

$$x \in \begin{cases} \text{Lớp I, nếu } f(x) = +1 \\ \text{Lớp II, nếu } f(x) = -1 \end{cases}$$

Có một loại hàm Kernel, việc xác định sử dụng hàm Kernel nào phụ thuộc vào bài toán cụ thể và được xác định theo kinh nghiệm.

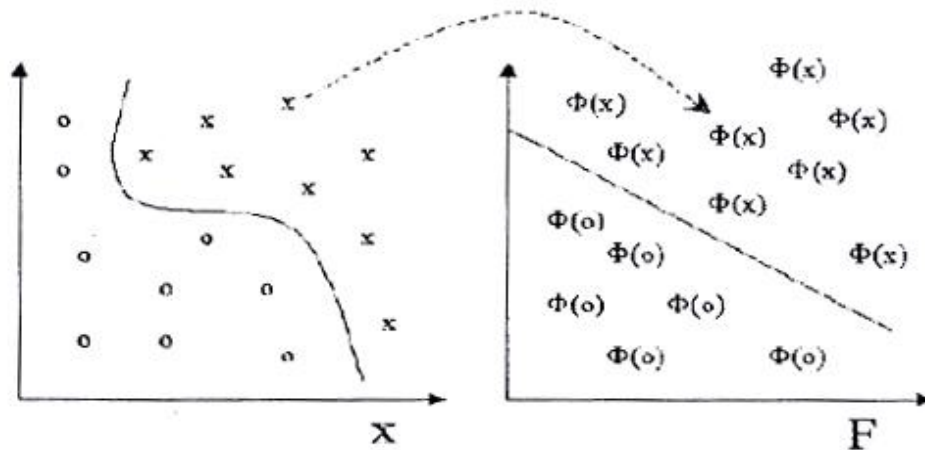
Sự phức tạp của hàm mục tiêu dẫn đến quá trình học phụ thuộc vào cách nó được diễn tả. Khi diễn tả dữ liệu một cách phù hợp, vấn đề học sẽ trở nên dễ dàng. Vì vậy, một việc làm rất phổ biến trong học máy là chuyển đổi dữ liệu từ không gian đầu vào X sang không gian đặc trưng:

$$x = (x_1, x_2, \dots, x_n) \mapsto \Phi(x) = (\Phi_1(x), \dots, \Phi_N(x))$$

Trong đó n là số chiều của đầu vào (số thuộc tính) và N là số chiều của không gian đặc trưng. Dữ liệu sẽ được chuyển vào không gian đặc trưng với $N > n$.

Không gian đặc trưng kí hiệu là F:

$$F = \{\Phi(x) | x \in X\}.$$



Hình 2.7 Ánh xạ Φ từ không gian dữ liệu X sang không gian đặc trưng F

Hình trên là ví dụ về ánh xạ từ không gian đầu vào hai chiều sang không gian đặc trưng hai chiều, trong đó dữ liệu không thể phân chia bởi hàm tuyến tính trong không gian đầu vào, nhưng có thể phân chia tuyến tính khi chuyển sang không gian đặc trưng.

2.2.2 Cơ sở lý thuyết SVM

Ý tưởng: Cho trước một tập huấn luyện, được biểu diễn trong không gian vector, trong đó mỗi tài liệu là một điểm, phương pháp này tìm ra một siêu phẳng f quyết định tốt nhất có thể chia các điểm trên không gian này thành hai lớp riêng biệt tương ứng là lớp + và lớp -. Chất lượng của siêu phẳng này được quyết định bởi khoảng cách (gọi là biên) của điểm dữ liệu gần nhất của mỗi lớp đến mặt phẳng này. Khi đó, khoảng cách biên càng lớn thì mặt phẳng quyết định càng tốt, đồng thời việc phân loại càng chính xác.

2.2.2.1. Bài toán phân 2 lớp với SVM

Bài toán đặt ra là: Xác định hàm phân lớp để phân lớp các mẫu trong tương lai, nghĩa là với một dữ liệu mới x_t thì cần phải xác định x_t được phân vào lớp +1 hay lớp -1?.

Nếu dữ liệu phân chia tuyến tính, hàm quyết định được xác định:

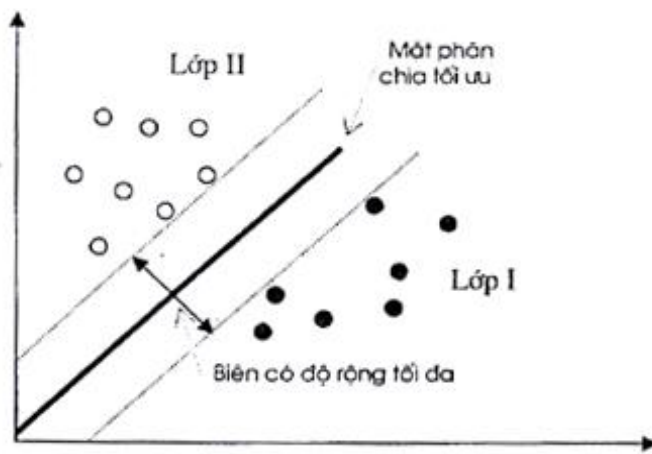
$$f(x) = w^T x + b \tag{2.4}$$

Với w là vector l chiều và b là vector vô hướng.

Siêu phẳng phân chia thỏa:

$$y_i(w^T x_i + b) \geq 1 \text{ với } i = 1, \dots, l \tag{2.5}$$

Siêu phẳng phân chia tối ưu là siêu phẳng chia hai lớp với khoảng cách giữa hai lớp là lớn nhất.



Trong đó:

- : các điểm có nhãn +1 (Lớp I)
- : các điểm có nhãn -1 (Lớp II)

Hình 2.8. Minh họa cho bài toán phân hai lớp

Siêu phẳng tối ưu có thể xác định được bằng việc giải bài toán tối ưu bậc hai:

$$\text{cực tiểu } \frac{1}{2} \|w\|^2 \tag{2.6}$$

Với ràng buộc $y_i(w \cdot x_i + b) \geq 1$

Nếu số thuộc tính của các mẫu dữ liệu là lớn, chúng ta có thể đơn giản hóa phép tính bằng cách chuyển bài toán với điều kiện Kuhn-Tucker tương đương với phương trình Lagrange kép. Phương trình Lagrange cho (2.6) được mô tả như sau:

$$L(w, b, \alpha) = \frac{1}{2}(w \cdot w) - \sum \alpha_i [y_i((w_i \cdot x_i) + b) - 1] \tag{2.7}$$

Với $\alpha = (\alpha_1, \dots, \alpha_M)$ là bội số Lagrange.

Phương trình kép trên trở thành : *cực đại* $L(w, b, \alpha)$ (2.8)

Với ràng buộc $\alpha_i \geq 0, i = 1, \dots, M$

Vì phân từng phần của 3.4 lần lượt với w và b ta được:

$$\frac{\partial L}{\partial w}(w, b, \alpha) = w - \sum_{i=1}^M y_i \alpha_i x_i = 0$$

$$\frac{\partial L}{\partial w}(w, b, \alpha) = \sum_{i=1}^M y_i \alpha_i = 0 \tag{2.9}$$

Từ (2.7), (2.8) và (2.9) ta có bài toán tối ưu sau:

$$\text{cực đại } w(\alpha) = \sum_{i=1}^M \alpha_i - \frac{1}{2} \sum_{i,k=0}^M \alpha_i \alpha_k y_i y_k x_i x_k \quad (2.10)$$

$$\text{Với ràng buộc } \sum_{i=1}^M y_i \alpha_i = 0, \alpha_i \geq 0, i = 1, \dots, M$$

Số lượng biến trong bài toán tối ưu trên chính là số mẫu trong tập dữ liệu huấn luyện.

Giả sử ta tìm được cặp nghiệm tối ưu trên chính là số mẫu trong tập dữ liệu huấn luyện.

Giả sử ta tìm được cặp nghiệm tối ưu là α^* và b^* . Theo lý thuyết Karush-Kuhn-Tucker, điều kiện xảy ra đẳng thức (2.5) tương ứng với mỗi cặp huấn luyện vào-ra (x_i, y_i) khi và chỉ khi α_i^* khác 0. Trong trường hợp này, mẫu huấn luyện chính là một support vector. Giải (2.10) với $\alpha = (\alpha_1, \dots, \alpha_M)$ ta có thể xác định được các support vector cho lớp I và lớp II. Do siêu phẳng tối ưu nằm trong số các support vector vừa tìm được của lớp I và lớp II, nên b^* được tính bởi:

$$b^* = -\frac{1}{2} \sum_{k=-1}^M y_k \alpha_k^* (s_1 x_k + s_2 x_k)$$

Với s_1 và s_2 lần lượt là các suport vector của lớp I và lớp II.

Trong trường hợp các mẫu huấn luyện không thể phân chia tuyến tính, chúng ta đưa thêm một biến phụ không âm ξ_i vào (2.5), cộng thêm vào hàm (2.8), và tổng của các biến phụ được nhân với tham số C. Điều này tương ứng với việc cộng biên trên của C với α . Trong cả hai trường hợp trên, hàm ra quyết định là như nhau và được tính bằng:

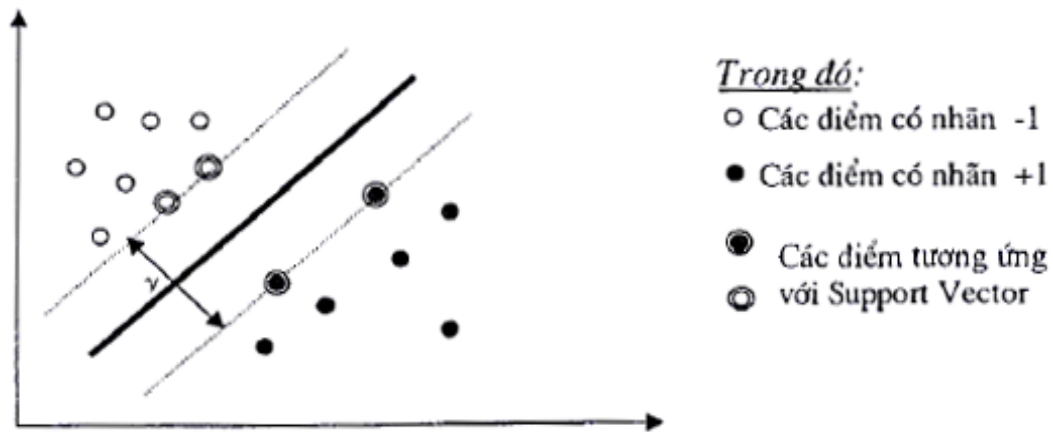
$$f(x) = \sum_{i=1}^M \alpha_i^* y_i x_i \cdot x + b^*$$

Với mỗi một mẫu dữ liệu ban đầu, nó sẽ được phân lớp như sau:

$$x \in \begin{cases} \text{Lớp I, nếu } f(x) > 0 \\ \text{Lớp II, nếu ngược lại} \end{cases}$$

Phương pháp SVM cho bài toán phân lớp được mô tả thông qua ví dụ sau:

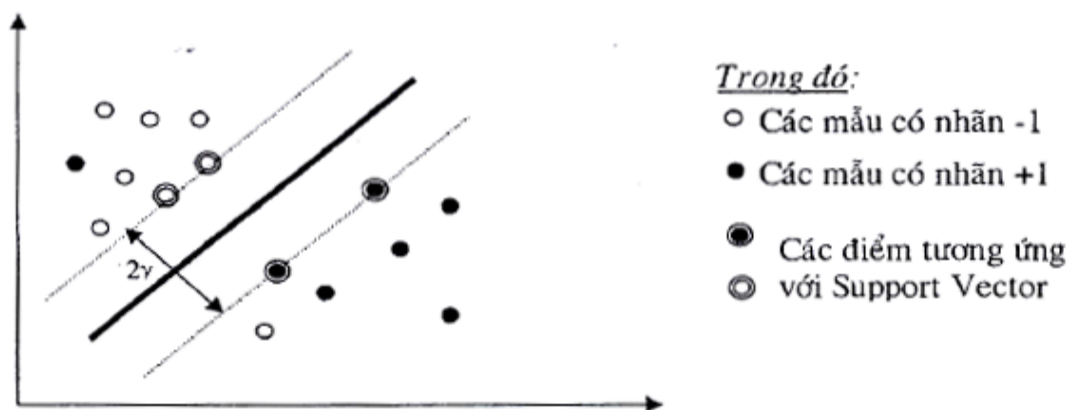
Xét không gian 2 chiều (n=2), các điểm của tập mẫu được cho như hình 2.5:



Hình 2.9. Minh họa bài toán phân hai lớp với phương pháp SVM

Để xác định hàm phân lớp dựa trên phương pháp SVM, ta sẽ tiến hành tìm 2 siêu phẳng song song (tương ứng với 2 đường nét đứt trong hình 2.5 ở trên) sao cho khoảng cách γ giữa chúng là lớn nhất có thể để phân tách hai lớp này ra làm hai phía. Hàm phân tách tương ứng với phương trình siêu phẳng nằm giữa hai siêu phẳng tìm được (đường nét đậm).

Hình 2.5 thể hiện trường hợp có thể tìm được siêu phẳng phân tách, dữ liệu trong trường hợp này gọi là phân tách tuyến tính. Bây giờ ta xét trường hợp dữ liệu không gian phân tách tuyến tính nó được minh họa trong hình 2.6 như sau:



Hình 2.10. Bài toán SVM trong trường hợp dữ liệu mẫu không phân tách tuyến tính

Ta thấy rằng trong hình 2.6 ở trên có những mẫu mặc dù có nhãn +1, nhưng lại bị “rơi” vào phía các mẫu có nhãn -1 và ngược lại.

Trong trường hợp này, phương pháp SVM sẽ sử dụng một phép ánh xạ dữ liệu mẫu vào không gian mới có số chiều lớn hơn sao cho tập mẫu này sẽ là phân tách tuyến tính trong không gian đó (ta gọi không gian mới này là không gian đặc trưng).

Trong không gian đặc trưng này, ta vẫn tiến hành tìm khoảng cách cực đại giữa hai siêu phẳng song song để phân cách dữ liệu mẫu.

Các điểm mà nằm trên hai siêu phẳng phân cách được gọi là support vector. Các điểm này sẽ quyết định đến hàm phân cách dữ liệu.

Trong thực tế, để thuận tiện cho quá trình tính toán, dữ liệu mẫu được ánh xạ vào không gian đặc trưng bằng cách sử dụng các hàm hạt nhân, cách làm này sẽ làm tăng tốc độ tính toán và đảm bảo rằng dữ liệu sẽ gần như được phân tách tuyến tính.

Điều kiện tối ưu Karush-Kuhn-Tucker (KKT)

Điều kiện KKT có vai trò quan trọng đối với bài toán tối ưu ràng buộc. Với bài toán trên, điều kiện KKT có thể phát biểu:

$$\frac{\partial L}{\partial w}(w, b, \alpha) = w_v - \sum_i y_i \alpha_i x_{iv} = 0 \quad (2.11)$$

$$\frac{\partial L}{\partial b}(w, b, \alpha) = -\sum_i y_i \alpha_i = 0 \quad (2.12)$$

$$y_i(x_i w + b) - 1 \geq 0 \quad i = 1, \dots, l \quad (2.13)$$

$$\alpha_i(y_i(w x_i + b) - 1) = 0 \quad \forall i \quad (2.14)$$

Điều kiện KKT là điều kiện cần và đủ để w, b, α là một lời giải (Fletcher, 1987). Do đó giải bài toán SVM tương đương với việc tìm lời giải cho các điều kiện KKT.

Vector w được tính bằng cách huấn luyện, nhưng b thì không. Tuy nhiên b dễ dàng tính được khi sử dụng các điều kiện KKT bổ sung, công thức (2.14), bằng việc chọn i có $\alpha_i \neq 0$ và tính b (lấy giá trị trung bình của b từ các phương trình trên).

2.2.2.2. Bài toán phân nhiều lớp với SVM

Để phân nhiều lớp thì kỹ thuật SVM nguyên thủy sẽ chia không gian dữ liệu thành 2 phần và quá trình này lặp lại nhiều lần. Khi đó hàm quyết định phân dữ liệu vào lớp thứ I của tập n 2-lớp sẽ là:

$$f_i(x) = w_i^l x + b_i$$

Những phần tử x là support vector sẽ thỏa điều kiện:

$$f_i(x) = \begin{cases} 1, & \text{nếu thuộc lớp } i \\ -1, & \text{nếu thuộc phân còn lại} \end{cases}$$

Như vậy, bài toán phân nhiều lớp sử dụng phương pháp SVM hoàn toàn có thể thực hiện giống như bài toán hai lớp. Bằng cách sử dụng chiến lược “một đối một” (one – against – one).

Giả sử bài toán cần phân loại có k lớp ($k > 1$), chiến lược “một đối một” sẽ tiến hành $k(k-1)/2$ lần phân lớp nhị phân sử dụng phương pháp SVM. Mỗi lớp sẽ tiến hành phân tách với $k-1$ lớp còn lại để xác định $k-1$ hàm phân tách dựa vào bài toán phân hai lớp bằng phương pháp SVM.

Ứng dụng của bài toán phân nhiều lớp trong xử lý ảnh chính là bài toán nhận diện mặt người và phát hiện trạng thái mắt của mặt người trong ảnh.

2.2.2.3. Trường hợp dữ liệu không thể phân tách được

○ Nhận xét:

- ✓ Trên thực tế có rất nhiều trường hợp dữ liệu mẫu bị nhiễu, nói chung là không có sự phân tách tuyến tính trong không gian đặc trưng.
- ✓ Trong trường hợp vượt quá giới hạn, các hệ số Lagrange sẽ rất lớn, không nhận được lời giải khả thi.
- ✓ Một cách “làm sạch” dữ liệu là xếp hạng các dữ liệu mẫu tùy nhiên lại khó cho việc phân lớp.

Thuật toán trên chỉ phù hợp cho trường hợp tập dữ liệu học có thể phân cách được, khi áp dụng cho dữ liệu không thể phân cách tuyến tính, sẽ không nhận được lời giải khả thi do hàm **Lagrange** lớn. Để mở rộng ý tưởng này cho trường hợp dữ liệu không thể phân cách ta đưa thêm các “biến mềm” slack (slack variables) dương $\xi_i \geq 0, i = 1, \dots, l$ cho bài toán tối ưu:

$$\min_{w,b} \frac{1}{2} (w \cdot w)$$

$$\text{với } y_i((x_i w) + b) \geq 1 - \xi_i, \quad i = 1, \dots, l$$

Khi có lỗi xuất hiện, ξ_i tương ứng sẽ lớn hơn 1, như vậy $\sum_i \xi_i$ là biên trên của số lỗi huấn luyện. Do đó một cách để gán thêm lỗi huấn luyện, là thay đổi hàm mục tiêu từ việc cực tiểu $(w \cdot w)/2$ sang:

$$\min_{w,b} (w \cdot w)/2 + C \sum_{i=1}^l \xi_i$$

$$\text{với } y_i((x_i w) + b) \geq 1 - \xi_i, \quad i = 1, \dots, l$$

Trong đó C là tham số do người dùng chọn, và C càng lớn thì tỉ lệ lỗi sẽ càng thấp. Trường hợp này ta gọi là bài toán **dạng C-SVC**.

Tương ứng dạng Lagrange cho vấn đề tối ưu khoảng cách mềm chuẩn 1 là:

$$L(w, b, \xi, \alpha, r) = \frac{1}{2}(w \cdot w) + C \sum_{i=1}^l \xi_i - \sum_{i=1}^l \alpha_i [y_i(\langle x_i, w \rangle + b) - 1 + \xi_i] - \sum_{i=1}^l r_i \xi_i$$

với $\alpha_i, r_i \geq 0$.

Để xác định w, b và ξ_i ta dùng:

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^l y_i \alpha_i x_i = 0$$

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - r_i = 0$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^l y_i \alpha_i = 0$$

Khi đó bài toán trở thành:

Tìm α để cực đại hóa biểu thức:

$$\sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i y_i \alpha_j y_j K(x_i, x_j)$$

Với sự khác nhau duy nhất là ràng buộc $C - \alpha_i - r_i = 0$, cùng với $r_i \geq 0$, dẫn đến $\alpha_i < C$, trong khi $\xi_i \neq 0$ nếu $r_i = 0$ và vì vậy $\alpha_i = C$.

Các điều kiện KKT được bổ sung cho $\alpha_i^* \neq 0$:

$$y_i(\langle x_i, w^* \rangle + b^*) - 1 + \xi_i^* = 0 \tag{2.15}$$

Và khi $\xi_i^* \neq 0$ ta có:

$$\alpha_i^* - C = 0$$

Điều này dẫn đến rằng các biến slack sẽ khác không chỉ xảy ra khi $\alpha_i^* = C$ (điều đó cũng nói rằng mẫu i là không phân lớp được). Các điểm với các biến slack khác không có khoảng cách hình học nhỏ hơn $1/\|w^*\|$. Các điểm có $0 < \alpha_i^* < C$ nằm tại chỗ có khoảng cách $1/\|w^*\|$ tính từ siêu phẳng. Do khi $\alpha_i^* = C, \xi_i = 0$. Có

thể duy nhất tính toán hệ số b^* từ các điểm tương ứng với $0 < \alpha_i^* < C$ sử dụng điều kiện $y_i((x_i w^*) + b^*) - 1 = 0$, suy ra:

$$b^* = \frac{1}{y_i} - \sum_{j=1}^l \alpha_j^* y_j K(x_i, x_j) \text{ với } 0 < \alpha_i^* < C$$

Thêm nữa khoảng cách được tính như sau:

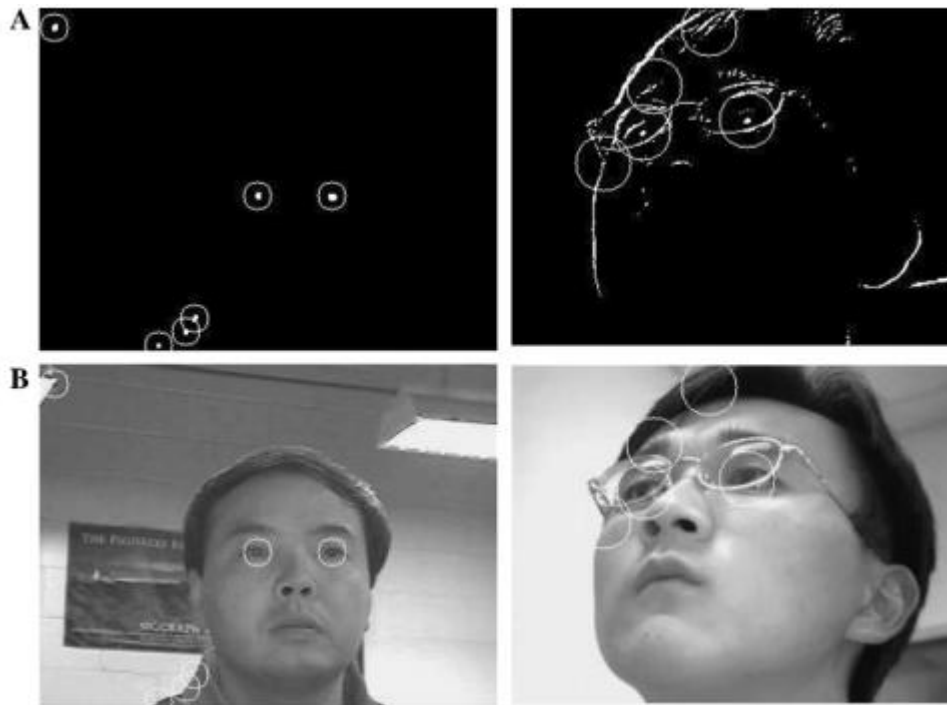
$$\gamma = \left(\sum_{i,j \in SV} \alpha_i^* y_i \alpha_j^* y_j K(x_i, x_j) \right)^{-1/2}$$

2.2.3. Huấn luyện SVM

Sử dụng SVM, dữ liệu huấn luyện là cần thiết để có được tối ưu siêu máy bay. Một hình ảnh mắt được biểu diễn như là một vector tối bao gồm các giá trị pixel. Dự án này, sau khi có được vị trí của các ứng cử viên học sinh bằng cách sử dụng phương pháp đề cập ở trên, chúng ta có được những tiểu hình ảnh từ hình ảnh tối theo những tích chức như inFig hiển thị 9. Thông thường, đôi mắt được bao gồm trong những hình ảnh cắt của 20,20 điểm ảnh. Các dữ liệu hình ảnh cắt được chế biến sử dụng biểu đồ cân bằng và bình thường hóa với một loạt trước khi đào tạo [0,1].

Những hình ảnh đào tạo mắt được chia thành hai bộ: bộ tích cực và thiết lập tiêu cực. Trong các thiết lập hình ảnh tích cực, chúng tôi bao gồm hình ảnh mắt của cái nhìn khác nhau, mức độ khác nhau của mở cửa, khuôn mặt khác nhau gây ra, đôi tượng khác nhau, và (có / không) có kính. Các hình ảnh không mắt được đặt trong set.Figs hình ảnh tiêu cực 10 và ví dụ 11 contain mắt và hình ảnh không mắt trong các bộ tương ứng. Sau khi hoàn thành các bước trên, chúng ta có được một tập huấn luyện, trong đó có 558 hình ảnh tích cực và 560 hình ảnh tiêu cực. Để có được độ chính xác tốt nhất, chúng ta cần phải xác định thông số tốt nhất cho SVM.

Trong bảng 1, chúng tôi danh sách ba hạt nhân SVM khác nhau với các thiết lập thông số khác nhau và mỗi SVM đã được thử nghiệm trên 1.757 hình ảnh ứng cử viên mắt thu được từ những người khác nhau. Từ bảng trên, chúng ta có thể thấy rằng sự chính xác tốt nhất chúng tôi có thể đạt được là 95,5037%, sử dụng một hạt nhân Gaussian với aróf 3.



Hình 2.11 (A) hình ảnh khác biệt thresholded chồng với ứng viên học sinh có thể. (B) hình ảnh tối được đánh dấu với các ứng cử viên mắt có thể theo vị trí của các ứng cử viên học sinh(A)

2.2.4. Phát hiện trạng thái mắt với SVM

Trong khi phát hiện mắt, chúng tôi cắt các khu vực trong hình ảnh học sinh tối theo vị trí của các ứng cử viên học sinh trong các hình ảnh khác nhau như inFig hiển thị. 9B. Sau khi một số tiền xử lý trên những hình ảnh ứng cử viên mắt, họ sẽ được cung cấp cho đào tạo SVM phân loại. Các SVM được đào tạo sẽ phân loại các vector vào lớp mắt đầu vào hoặc lớp.Fig không mắt. 12 cho thấy các phân loại SVM mắt một cách chính xác xác định các khu vực mắt thực sự như đánh dấu.

Xác minh học sinh với SVM hoạt động khá tốt và có thể khái quát cho người dân của cùng một chủng tộc. Tuy nhiên, cho những người từ một cuộc đua mà là khác nhau đáng kể từ những hình ảnh trong đào tạo, SVM có thể thất bại và cần phải được đào tạo lại. SVM có thể làm việc điều kiện chiếu sáng khác nhau do bình thường cường độ cho hình ảnh đào tạo thông qua cân bằng biểu đồ.

Chương 3

KẾT QUẢ THỰC NGHIỆM

Trong phần này, chúng tôi sẽ trình bày kết quả từ một thử nghiệm rộng rãi, chúng tôi tiến hành để xác nhận việc thực hiện theo dõi mắt tích hợp của chúng tôi trong các điều kiện khác nhau.

3.1 Bài toán

Chương trình thử nghiệm sử dụng kỹ thuật SVM để thực hiện đọc video đầu vào và thực hiện phát hiện trạng thái mắt của mặt người trong ảnh.

Đầu vào: một video các trạng thái của mắt.

Đầu ra: video đã được phát hiện trạng thái mắt của mặt người trong ảnh

Cụ thể đây là một chương trình phát hiện mắt của mặt người trong ảnh được viết trên nền visual C[#] của Microsoft, sử dụng thư viện mã nguồn mở OpenCV.

Mắt thường một phần hoặc hoàn toàn occluded bằng mặt do sự định hướng mặt nghiêng hoặc bằng tay hoặc các đối tượng khác. Một đôi mắt tốt sẽ có thể theo dõi mắt theo một phần tắc và có thể phát hiện tắc hoàn toàn và tái phát hiện mắt sau khi tắc hoàn toàn được loại bỏ. InFig. 26, hai người đang di chuyển ở phía trước của máy ảnh, và một người tôi? Đôi mắt được occluded bởi một cái đầu nữa không? Khi họ đang băng qua. Như thể hiện inFig. 26, khi người di chuyển phía sau từ phải sang trái, người đứng đầu của người phía trước bắt đầu hấp thụ mắt, bắt đầu với một và sau đó hai mắt nhận được hoàn toàn occluded. Như thể hiện, theo dõi của chúng tôi vẫn có thể theo dõi một cách chính xác một mắt ngay cả khi nó được occluded một phần. Khi cả hai mắt hoàn toàn occluded, theo dõi của chúng tôi phát hiện tình trạng này. Ngay sau khi xuất hiện trở lại đôi mắt trong hình ảnh, theo dõi mắt của chúng tôi sẽ nắm bắt được đôi mắt từng người một ngay lập tức như hình. 26. Thử nghiệm này cho thấy sự vững mạnh của phương pháp của chúng tôi để occlusions.



Hình 3.1 Sự di chuyển ánh mắt của mặt người trước ảnh

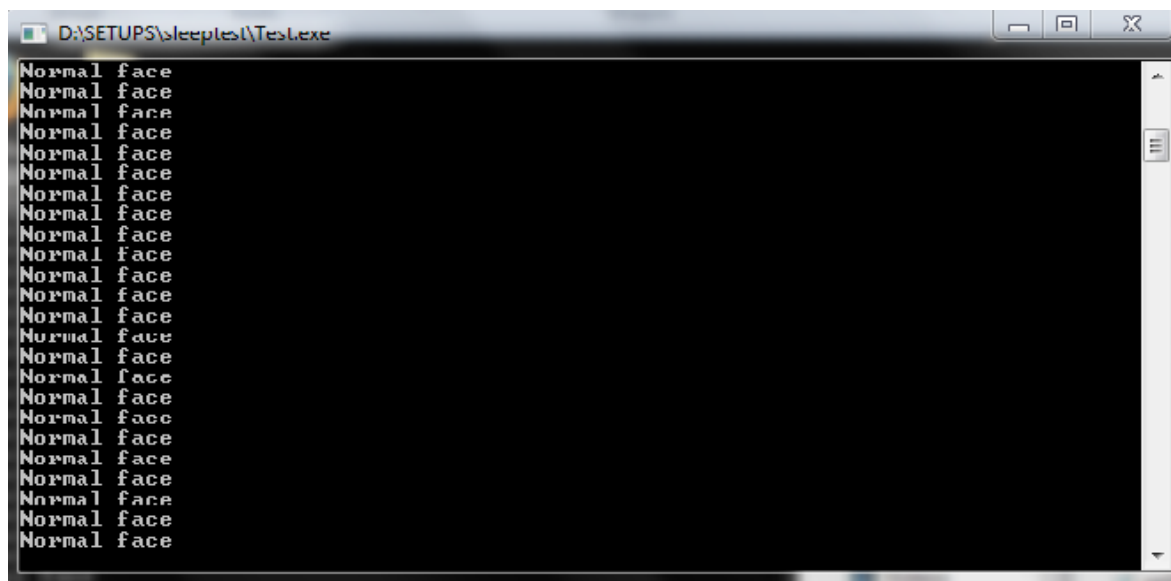
3.2. Cách bước thực hiện

Bước 1: Đưa vào chương trình video cần phát hiện mắt của mặt người.

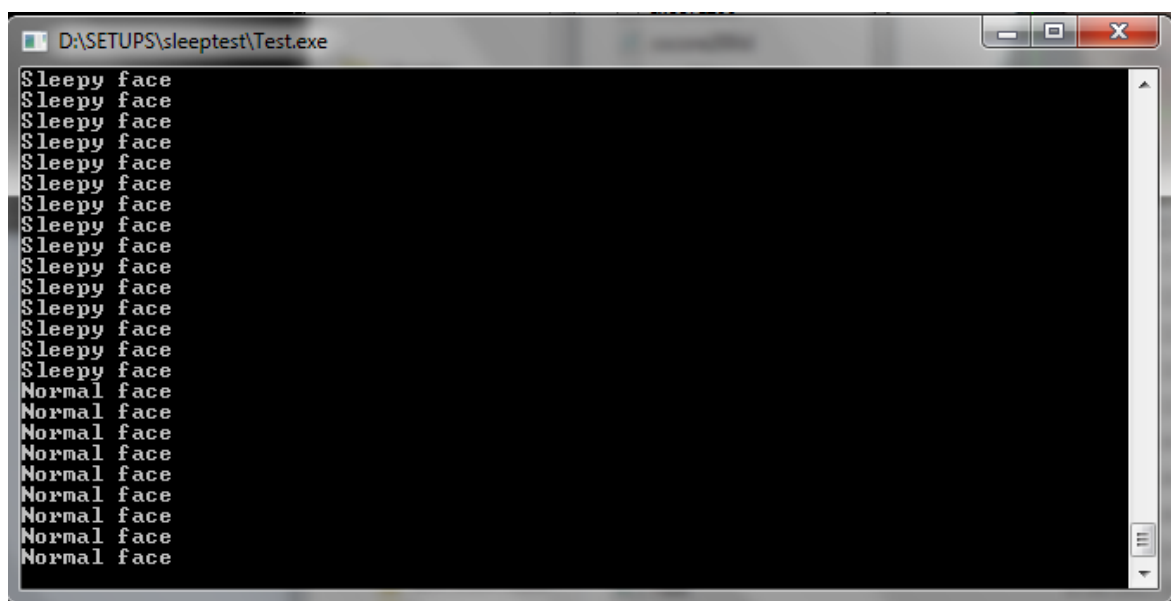
Bước 2: Thực hiện trích chọn đặc trưng trên video.

Bước 3: SVM quyết định mắt của mặt người trong ảnh.

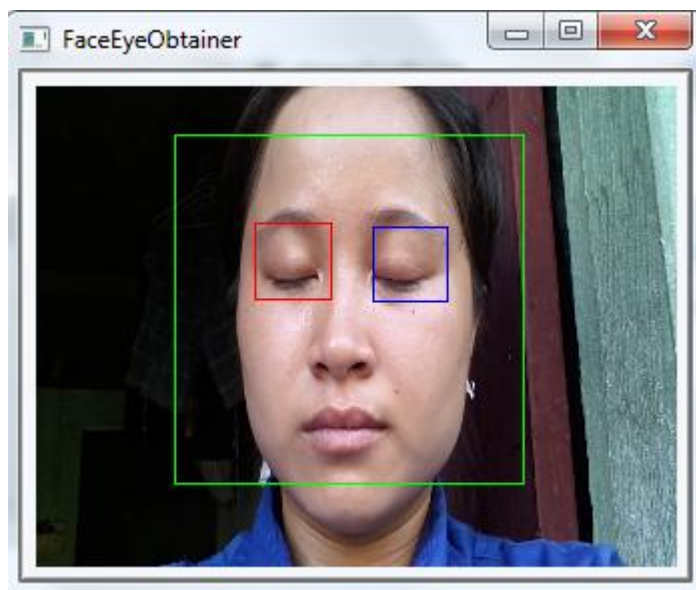
3.3. Giao diện chương trình và kết quả thực nghiệm



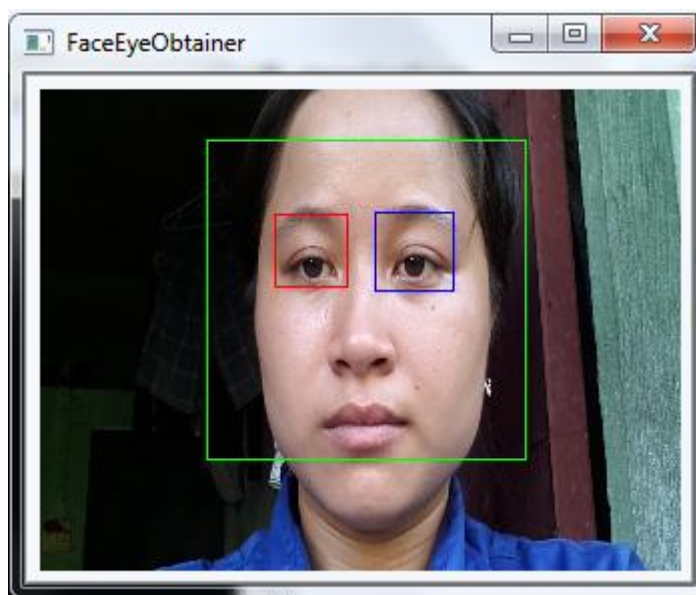
Hình 3.2 Ảnh chương trình đang chạy quá trình đầu



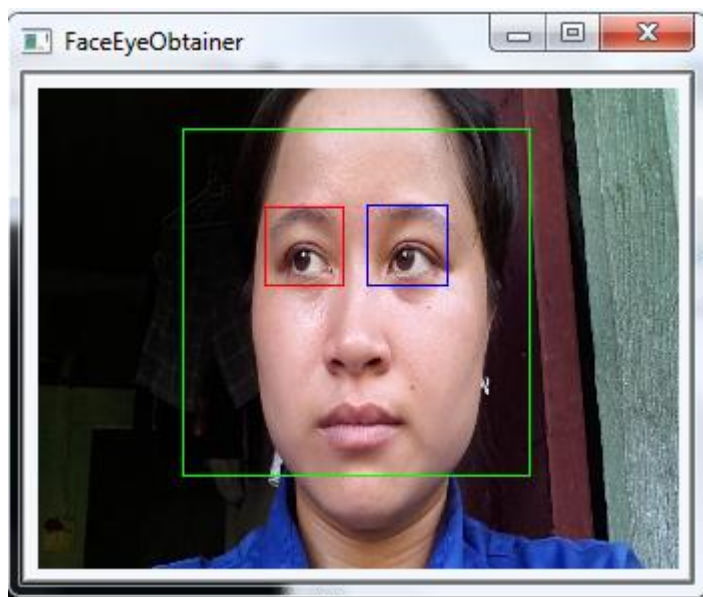
Hình 3.3 Quá trình tiếp theo



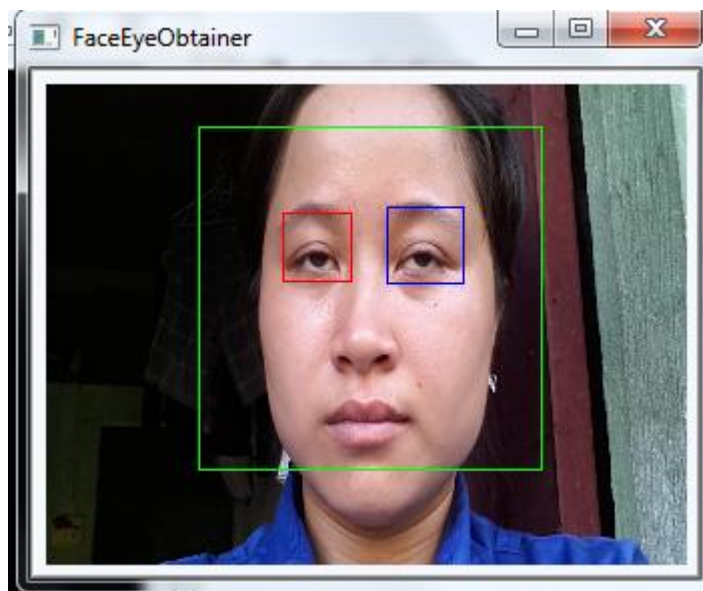
Hình 3.4 Trạng thái mắt nhắm



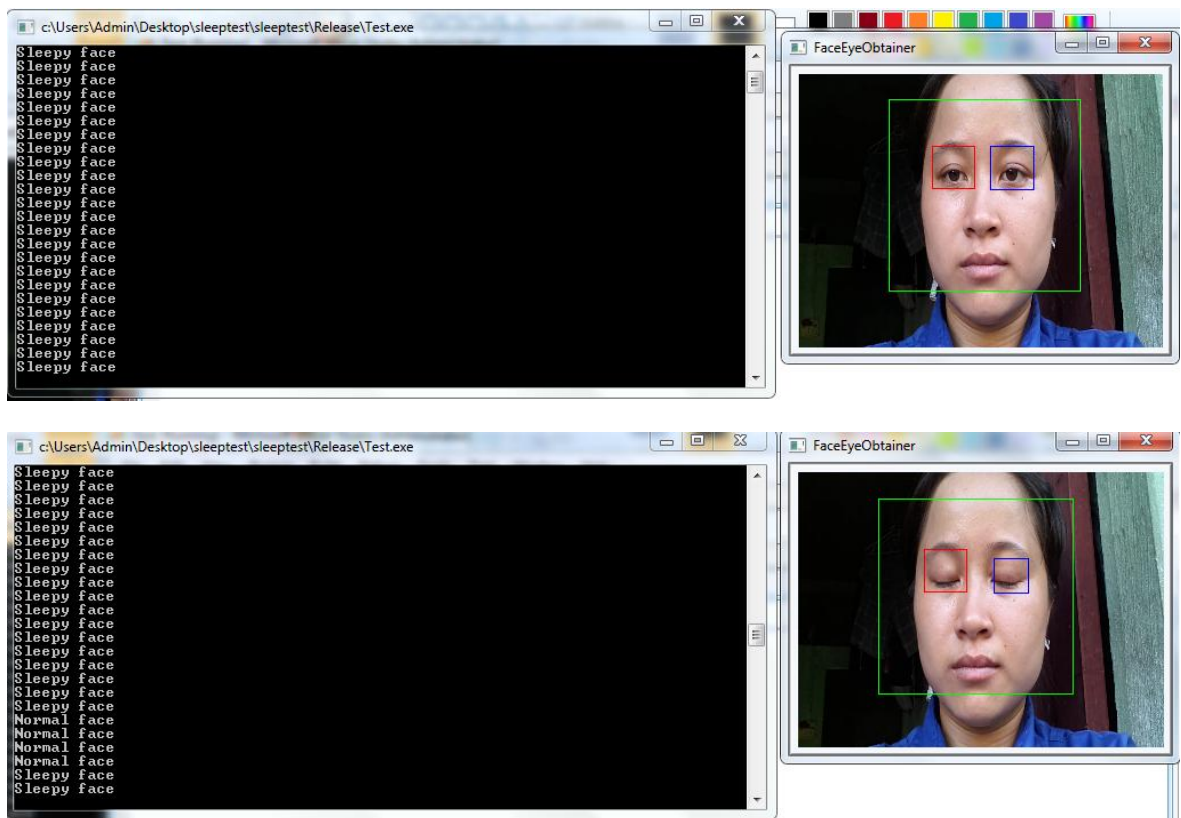
Hình 3.5 Trạng thái mắt mở nhìn thẳng



Hình 3.6 Trạng thái mắt mở nhìn sang bên



Hình 3.7 Trạng thái mắt nhìn buồn ngủ



Hình 3.8 Chương trình chạy và kết quả

KẾT LUẬN

Bài toán phát hiện mắt của mặt người đã được đặt ra từ lâu và đi cùng với nó là hàng loạt các đề tài khoa học được xác lập, các công trình nghiên cứu được công bố, các ứng dụng được triển khai. Bài toán này luôn đặt ra nhiều thách thức vì những khó khăn của nó. Các khó khăn chứng tỏ rằng bất cứ phương pháp giải quyết bài toán xác định mắt của mặt người sẽ không thể tránh khỏi một số khiếm khuyết nhất định. Mỗi hướng tiếp cận được đưa ra đều đã đạt được những thành quả nhất định, hướng nào cũng có những thành công, những hạn chế vì bài toán này là bài toán không có lời giải tối ưu cho mọi trường hợp. Tuy nhiên, do tính cấp thiết từ yêu cầu của thực tế mà đây luôn là một đề tài hấp dẫn các nhà khoa học, các chuyên gia nghiên cứu và ứng dụng. Đặc biệt là ở Việt Nam, nơi mà các hệ thống nhận dạng, các hệ thống giám sát chưa được phát triển mạnh. Những ứng dụng vẫn chủ yếu là các phần mềm đi kèm với các thiết bị chuyên dụng khá đắt tiền.

Trong luận văn này, tác giả đã tập trung nghiên cứu hướng phát hiện trạng thái mắt của mặt người trong ảnh, cụ thể đã đạt được những kết quả sau :

- Trình bày khái quát về xử lý ảnh và bài toán phát hiện trạng thái mắt của mặt người trong ảnh.
- Hệ thống hoá một số kỹ thuật phát hiện mắt của mặt người trong ảnh
- Cài đặt thử nghiệm thuật toán phát hiện mắt của mặt người trong ảnh.
 - Đầu tiên, việc sử dụng của SVM để phát hiện học sinh bổ sung với đôi mắt phát hiện dựa trên học sinh sáng từ ánh sáng hồng ngoại, cho phép phát hiện đôi mắt trong sự hiện diện của các đối tượng tươi sáng khác,
 - Thứ hai, hai kênh (mắt sang học sinh và hình ảnh mắt sáng học sinh) được sử dụng để mô tả sự phân bố thống kê của mắt, trên cơ sở đó một ca dôi mắt trung bình được phát triển.
 - Thứ ba, mô hình mắt được cập nhật liên tục bằng mắt phát hiện thành công từ Kalman theo dõi cuối cùng để tránh tuyên truyền lỗi với theo dõi sự thay đổi có ý nghĩa.
 - Cuối cùng, quyết tâm thực nghiệm của kích thước cửa sổ tối ưu và mức lượng tử hóa cho sự thay đổi có nghĩa là theo dõi thêm vi-hances việc thực hiện các kỹ thuật của chúng tôi. Thí nghiệm cho thấy các cải tiến này đã dẫn đến sự cải thiện đáng kể trong mắt

manh mẽ theo dõi và chính xác hơn trackers mắt hiện có, đặc biệt là trong các điều kiện khác nhau được xác định

Nhận xét và hướng phát triển tương lai:

- Ưu điểm:
 - Xét về độ tin cậy trong xử lý, SVM là một thuật toán mà bản thân nó có khả năng xử lý tổng quát cao.
 - Về thuật toán của SVM dễ hiểu có tính thực tiễn cao trong nhiều lĩnh vực.
 - SVM là thuật toán có khả năng phân loại chính xác với tỉ lệ cao, thích hợp cho những bài toán phức tạp như nhận dạng mắt của mặt người trong ảnh.
- Nhược điểm:
 - Khiếm khuyết lớn nhất của thuật toán SVM đó là chỉ có khả năng phân biệt tập học nhị phân, vì lẽ đó nhiều ứng dụng trong thực tế sử dụng SVM là không hợp lý. Để đáp ứng cho phân loại nhiều lớp ta phải có nhiều cách tổ chức phức tạp dẫn đến chi phí tính toán cao.
 - Bài toán SVM và việc xử lý chủ yếu dựa trên nền tảng các mẫu dữ liệu gọi là vector hỗ trợ tìm được trước khi thực hiện tìm siêu mặt, nếu số vector hỗ trợ quá lớn thì dẫn đến chi phí cho thời gian tìm siêu mặt cũng như chi phí do giai đoạn nhận dạng là rất lớn. Vì vậy SVM thường ít thành công trong việc xử lý bài toán với khối lượng dữ liệu tập học lớn.
- Hướng phát triển:

Dò tìm trạng thái mắt của mặt người trong ảnh là một bước trong quá trình nhận dạng trạng thái mắt, một bài toán có tính thực tiễn cao, từ nguyên lý về thuật toán của quá trình phát hiện trạng thái mắt của mặt người bằng SVM, chúng ta hoàn toàn có thể xây dựng được nguyên lý cho việc phát triển một hệ thống nhận dạng trạng thái mắt người bằng cách sử dụng SVM trong một bài toán phân nhiều lớp.

TÀI LIỆU THAM KHẢO

- [1]. Đỗ Năng Toàn, Phạm Việt Bình (2007), *Giáo trình xử lý ảnh*, Nhà xuất bản Đại học Thái Nguyên.
- [2]. Lương Mạnh Bá, Nguyễn Thanh Thủy (2007), *Nhập môn xử lý ảnh số*, Nhà xuất bản KHKT.
- [3]. Trần Phước Long, Nguyễn Văn Lượng (2003), *Nhận dạng người dựa vào thông tin khuôn mặt xuất hiện trên ảnh*, Luận văn tốt nghiệp cử nhân tin học. Luận văn tốt nghiệp cử nhân tin học khoa CNTT, ĐH KHTN TP HCM.
- [4]. Nguyễn Hải Châu(2009), *Phát hiện mặt người trong ảnh và ứng dụng*, Luận văn tốt nghiệp cử nhân công nghệ thông tin khoa CNTT, ĐH Công nghệ ĐH Quốc Gia Hà Nội.
- [5]. Phạm Anh Phương, Ngô Quốc Tạo, Lương Chi Mai, *Trích chọn đặc trưng wavelet Haar kết hợp với SVM cho việc nhận dạng chữ viết tay tiếng việt*, Bài báo khoa học, Viện Công Nghệ Thông Tin Hà Nội.
- [6]. Phạm Anh Phương, Ngô Quốc Tạo, Lương Chi Mai, *Trích chọn đặc trưng wavelet Haar kết hợp với SVM cho việc nhận dạng chữ viết tay tiếng việt*, Bài báo khoa học, Viện Công Nghệ Thông Tin Hà Nội.
- [7]. Ignas Kukenys, Brendan McCane (2008), *Support Vector Machines for Human Face Detection*, Christchurch New Zealand.
- [8]. [4] X. Xie, R. Sudhakar, H. Zhuang, On improving eye feature extraction using deformable templates, *Pattern Recognit.* 27 (1994) 791–799.
- [9] K.M. Lam, H. Yan, Locating and extracting the eye in human face images, *Pattern Recognit.* 29 (1996) 771–779.
- [10]. *Computer Vision and Image Understanding* 98 (2005) 124–154, *Robust real-time eye detection and tracking under variable lighting conditions and various face orientations.*