

**ĐẠI HỌC ĐÀ NẴNG
TRƯỜNG CĐ CÔNG NGHỆ THÔNG TIN**

**BÁO CÁO TỔNG KẾT
ĐỀ TÀI KHOA HỌC VÀ CÔNG NGHỆ
CẤP CƠ SỞ**

**NGHIÊN CỨU PHƯƠNG PHÁP XÁC ĐỊNH HƯỚNG
NHÌN CỦA KHUÔN MẶT NGƯỜI TRONG ẢNH**

Mã số: t2017-07-04

Xác nhận của cơ quan chủ trì đề tài

Chủ nhiệm đề tài

Đà Nẵng, 12/2017

MỤC LỤC

DANH MỤC HÌNH ẢNH	3
CÁC CHỮ VIẾT TẮT	6
MỞ ĐẦU	7
CHƯƠNG 1. TỔNG QUAN VỀ BÀI TOÁN NHẬN DẠNG MẶT NGƯỜI.....	10
1.1 GIỚI THIỆU CHUNG VỀ BÀI TOÁN NHẬN DẠNG MẶT NGƯỜI	10
1.2 NHỮNG ỨNG DỤNG LIÊN QUAN	11
1.3 PHÁT HIỆN KHUÔN MẶT VÀ NHỮNG KHÓ KHĂN CỦA BÀI TOÁN NHẬN DẠNG KHUÔN MẶT	13
1.3.1 Phát hiện mặt người.....	13
1.3.2 Những khó khăn và thách thức của bài toán nhận dạng khuôn mặt	14
1.4 BÀI TOÁN PHÁT HIỆN MẶT NGƯỜI TRONG ẢNH.....	15
1.4.1 Giới thiệu tổng quát bài toán.....	15
1.4.2 Các hướng tiếp cận liên quan đến nhận dạng khuôn mặt.....	16
1.4.3 Phát hiện mặt người trong ảnh dựa trên ngưỡng màu của da người và phương pháp trừ nền.....	19
1.4.4 Phát hiện mặt người trong ảnh dựa trên các đặc trưng Haar.....	22
1.4.5 Mô hình 3D của đầu người.....	34
CHƯƠNG 2. CÁC PHƯƠNG PHÁP XÁC ĐỊNH HƯỚNG MẶT NGƯỜI TRONG ẢNH	36
2.1 PHƯƠNG PHÁP XÁC ĐỊNH HƯỚNG MẶT NGƯỜI TRONG ẢNH	36
2.1.1 Xác định hướng mặt người dựa trên tâm của hai mắt.....	36

2.1.2	Xác định hướng mặt người dựa trên các đặc tính về mắt, mũi, miệng	37
2.2	XÁC ĐỊNH GÓC NHÌN CỦA KHUÔN MẶT DỰA TRÊN CÁC ĐIỂM MỐC TRÊN KHUÔN MẶT.	43
2.2.1	Phát hiện các điểm mốc (Land marks) trên khuôn mặt.....	43
2.2.2	Chuyển đổi ma trận xoay qua 3 góc Euler.....	45
	CHƯƠNG 3. THỰC NGHIỆM VÀ KẾT QUẢ	46
3.1	GIỚI THIỆU.....	46
3.2	DỮ LIỆU ĐẦU VÀO	46
3.2.1	Tư thế khuôn mặt người.....	46
3.2.2	Tập ảnh đánh giá.....	47
3.3	CHƯƠNG TRÌNH THỬ NGHIỆM.....	51
3.4	MỘT SỐ TRƯỜNG HỢP ĐẶC BIỆT	59
	KẾT LUẬN.....	63
	TÀI LIỆU THAM KHẢO	65

DANH MỤC HÌNH ẢNH

Số hình	Tên hình	Trang
1.1	Ứng dụng của nhận diện khuôn mặt	14
1.2	Phương pháp trừ nền với nền không đổi	22
1.3	Cách tính Integral Image	25
1.4	Cách tính nhanh tổng các điểm ảnh của vùng D trên ảnh	26
1.5	Cách tính tổng điểm ảnh vùng D với đặc trưng xoay 45°	27
1.6	Sử dụng đặc trưng cơ bản để xác định vùng mắt	28
1.7	Dựng thẳng, xoay các hình chữ nhật trong cửa sổ phát hiện	29
1.8	Một số mẫu dương (Positive Sample) dùng cho việc xác định khuôn mặt	30
1.9	Một số mẫu âm (Negative Sample) dùng cho việc xác định khuôn mặt	31
1.10	Chuỗi các nhận dạng theo tầng	33
1.11	Mô hình 3D của đầu người	38
2.1	Xác định góc nghiêng dựa trên tâm hai mắt	36
2.2	Kết quả sau khi phát hiện các đặc điểm	40
2.3	Các góc quay của khuôn mặt	40
2.4	Hệ tọa độ 3D của khuôn mặt	41
3.1	Khuôn mặt chuẩn và các tư thế thay đổi	48

3.2	Phương pháp tạo thư viện ảnh	50
3.3	Tư thế của khuôn mặt khi quay ngang	50
Số hình	Tên hình	Trang
3.4	Tư thế của khuôn mặt khi cúi xuống, ngẩng lên	51
3.5	Các tư thế của khuôn mặt khi ngẩng lên trên và quay từ phải qua trái	51
3.6	Các tư thế của khuôn mặt khi nhìn xuống từ trái qua phải	52
3.7	Không xác định được khuôn mặt	53
3.8	Xác định sai góc do xác định sai đường bao của khuôn mặt	53
3.9	Phương pháp dùng đặc trưng Haar không nhận diện đủ chi tiết trên khuôn mặt	54
3.10	Phương pháp xác định điểm mốc nhận dạng chính xác	54
3.11	Phương pháp nhận dạng điểm mốc ở góc quay lớn	55
3.12	Các điểm mốc trên khuôn mặt	56
3.13	Đường bao khuôn mặt và các bộ phận trên mặt	57
3.14	Góc xoay của khuôn mặt	57
3.15	Góc xoay của khuôn mặt biểu diễn hướng Vector	58
3.16	Hiệu suất phát hiện	59
3.17	Hiệu suất ước lượng	59
3.18	Sử dụng phương pháp đặc trưng Haar đối với	60

	ảnh người đeo kính	
3.19	Sử dụng phương pháp điểm mốc với ảnh người có đeo kính	61
3.20	Phương pháp đặc trưng Haar với ảnh đeo kính và có râu	61
3.21	Phương pháp điểm mốc với ảnh đeo kính và có râu	62
3.22	Phương pháp điểm mốc với ảnh chụp nhiều người	62

CÁC CHỮ VIẾT TẮT

ADC:	Analog to Digital
SAT:	Summed Area Table
RSAT:	Rotated Summed Area Table
HCI:	Human Computer Interaction
TP:	True Positive
FP:	False Positive
HCC:	Haar Cascade Classifier
HOG:	Histogram of Oriented Gradients

MỞ ĐẦU

1. Tính cấp thiết của đề tài

Thời gian gần đây có rất nhiều công trình nghiên cứu các phương pháp xác định khuôn mặt người. Các nghiên cứu này đi từ những bài toán đơn giản như ảnh trắng đen đến ảnh màu, ảnh chỉ có một khuôn mặt đến ảnh có nhiều khuôn mặt, ảnh chứa khuôn mặt ở tư thế thẳng đứng đến ảnh có chứa khuôn mặt ở tư thế nhìn nghiêng. Trong đó việc xác định khuôn mặt ở tư thế nhìn nghiêng gặp nhiều khó khăn trong việc nhận dạng.

Trong giao tiếp, con người thường sử dụng khả năng định hướng và chuyển động của đầu để diễn tả cảm xúc, nó như một hình thức giao tiếp phi ngôn ngữ. Đối với máy tính, việc xác định tư thế của đầu đòi hỏi phải thực hiện hàng loạt các biến đổi ảnh gốc dựa trên các điểm ảnh qua một khái niệm cao hơn dựa trên các hướng khác nhau. Và cũng như việc nhận dạng khuôn mặt trong ảnh, phương pháp xác định góc của hướng nhìn cũng phải bất biến với sự thay đổi của hàng loạt yếu tố trong bức ảnh. Các yếu tố này bao gồm các hiện tượng vật lý như biến dạng máy ảnh, các phép chiếu hình học, sự thay đổi của ánh sáng cũng như các yếu tố như phụ kiện, kính, ria mép...

Việc xác định góc quay của đầu người cũng có nhiều mức độ khác nhau, một số phương pháp chỉ xác định ở mức độ thấp như quay trái, quay phải. Ở mức độ cao hơn, việc xác định hướng nhìn của đầu người sử dụng các thuật toán để xác định tư thế đầu theo nhiều hướng như cúi đầu-nhìn lên, nhìn trái-phải, nhìn xiên..., thông qua việc xác định các đặc điểm trên khuôn mặt như mắt, mũi, miệng, tai.

Việc phát hiện các đặc điểm khuôn mặt có thể thông qua một số phương pháp như xác định dựa trên màu sắc, dựa trên sự xuất hiện các điểm ảnh trên khuôn mặt hoặc dựa trên các thống kê, mẫu có sẵn, bộ nhận dạng... Mỗi phương pháp khi áp dụng đều có những ưu điểm và hạn chế riêng.

Đó là lý do tôi quyết định chọn đề tài: “Nghiên cứu phương pháp hướng nhìn của khuôn mặt người trong ảnh”.

2. Mục đích nghiên cứu

Đề tài tập trung nghiên cứu các phương pháp xác định góc nhìn của khuôn mặt thông qua các bộ nhận dạng và tập dữ liệu được huấn luyện từ trước. Các phương pháp này sẽ được kiểm nghiệm và đánh giá kết quả dựa trên bộ dữ liệu chuẩn.

3. Đối tượng và phạm vi nghiên cứu

Nghiên cứu về các phương pháp nhận dạng mặt người và các chi tiết trên khuôn mặt

Nghiên cứu về phương pháp ước lượng góc quay của khuôn mặt dựa trên các đặc trưng chi tiết của khuôn mặt. Nghiên cứu hai phương pháp xác định dựa trên đặc trưng Haar và xác định dựa trên điểm mốc.

Cuối cùng là thực thi thuật toán, đánh giá kết quả và so sánh hai phương pháp.

4. Phương pháp nghiên cứu

Kết hợp nghiên cứu lý thuyết và mô phỏng.

Các bước tiến hành cụ thể như sau:

- Thu thập các tài liệu và thông tin liên quan đến đề tài.
- Phân tích, nghiên cứu nội dung chính.

- Sử dụng ngôn ngữ lập trình C++, thư viện OpenCV, thư viện Dlib viết chương trình thực thi thuật toán và đưa chương trình chạy trên PC.
- Đánh giá so sánh kết quả.
- Đề xuất các giải pháp mở rộng và nâng cấp đề tài.

5. Bố cục đề tài

Ngoài phần mở đầu, kết luận, tài liệu tham khảo và phụ lục, báo cáo được tổ chức thành 3 chương với kết cấu như sau:

Chương 1: Tổng quan về bài toán nhận dạng mặt người

Chương 2: Một số kĩ thuật xác định hướng mặt người trong ảnh

Chương 3: Thực nghiệm và kết quả.

CHƯƠNG 1

TỔNG QUAN VỀ BÀI TOÁN NHẬN DẠNG MẶT NGƯỜI

1.1 GIỚI THIỆU CHUNG VỀ BÀI TOÁN NHẬN DẠNG MẶT NGƯỜI

Trong những năm qua, có rất nhiều công trình nghiên cứu về bài toán nhận dạng mặt người. Các nghiên cứu đi từ những bài toán đơn giản như việc nhận dạng một người trong ảnh trắng đen, nền trơn và mở rộng lên những bài toán phức tạp hơn như nhận dạng người trong ảnh màu và có nhiều mặt người trong ảnh. Đến nay các bài toán về xác định khuôn mặt đã được mở rộng với nhiều hướng nghiên cứu như nhận dạng khuôn mặt, xác định các đặc điểm trên khuôn mặt, theo dõi khuôn mặt, xác định hướng nhìn hay nhận dạng cảm xúc mặt người...

Phát hiện khuôn mặt là phần đầu tiên của một hệ thống nhận dạng mặt người. Các hệ thống nhận dạng khuôn mặt được phát triển từ những năm 1980 tuy nhiên còn nhiều điểm hạn chế về luật nhận dạng nên phạm vi áp dụng thực tế chưa rộng. Ngày nay, có rất nhiều phương pháp được đưa ra để xác định khuôn mặt trong ảnh số và nhiều phương pháp có độ chính xác rất cao. Các bài toán về xác định khuôn mặt đã được áp dụng rất rộng rãi trong đời sống hiện nay.

Bài toán xác định mặt người là một kỹ thuật dựa trên thị giác máy tính (Computer Vision) để xác định vị trí và kích thước của các khuôn mặt người trong ảnh. Kỹ thuật này cho phép xác định vị trí của khuôn mặt, bỏ qua các chi tiết khác xung quanh như nhà cửa, cây cối, đồ vật.

1.2 NHỮNG ỨNG DỤNG LIÊN QUAN

Như đã trình bày ở trên, bài toán nhận dạng mặt người hiện đang được ứng dụng trong rất nhiều lĩnh vực, từ ứng dụng bình thường trong đời sống đến ứng dụng để đảm bảo an ninh, quốc phòng.

Các ứng dụng cơ bản có thể kể tới như:

- Camera chống trộm: Các hệ thống camera chống trộm sẽ xác định đâu là con người và theo dõi hành động của người đó xem có gì phạm pháp không như lấy trộm đồ, xâm phạm trái phép vào một khu vực nào đó.

- Bảo mật: Các ứng dụng về bảo mật rất đa dạng với nhiều loại công nghệ khác nhau sử dụng các đặc điểm sinh trắc học trong đó có nhận dạng khuôn mặt, công nghệ này sẽ chụp hình người truy cập và so khớp các đặc điểm của người đó. Các đặc điểm này được máy “học” thuộc và sử dụng cho các lần đăng nhập sau.

- Lưu trữ khuôn mặt: Xác định mặt người nhằm mục đích lưu trữ có thể được ứng dụng tại một số nơi như trạm rút tiền tự động ATM. Việc lưu trữ khuôn mặt người rút tiền sẽ giúp các ngân hàng quản lý việc rút tiền của khách hàng, khi có sự cố như thẻ AMT bị mất trộm hoặc làm giả thì việc giải quyết và xử lý sẽ dễ dàng hơn.

- Xác minh tội phạm: Dựa vào ảnh một người để xác định xem người ấy có phải là tội phạm không, hiện nay vấn đề an ninh, chống khủng bố đang là vấn đề cấp thiết. Các hệ thống xử lý ảnh có thể ghi hình nơi công cộng, truy vấn cơ sở dữ liệu và phát hiện đối tượng trong ảnh. Ứng dụng này giúp cơ quan an ninh quản lý con người tốt hơn.

Các ứng dụng khác:

- Điều khiển vào ra: ứng dụng cho các văn phòng, công ty, kết hợp với các phương pháp truyền thống như vân tay, thẻ từ, có thể kết hợp

thêm phương pháp nhận dạng khuôn mặt. Tránh việc giả mạo thẻ từ cũng như sử dụng vân tay giả.

-An ninh xuất nhập cảnh: Tại cái khu vực cửa khẩu, các cảng hàng không quốc tế, khi khách hàng làm thủ tục sẽ có hệ thống ghi hình, hệ thống này sẽ tự tìm kiếm và so khớp các thông tin của khách hàng. Giúp hạn chế việc tội phạm xuất nhập cảnh.

-Tìm kiếm dữ liệu: Việc nhận dạng khuôn mặt trong các video sẽ hỗ trợ đắc lực cho việc tìm kiếm. Ví dụ có thể tìm các video có hình ảnh của một nhân vật cụ thể nào đó.

-Xác định cảm xúc: Một số thiết bị hiện nay có thể phát hiện cảm xúc người sử dụng, biết được các trạng thái vui, buồn... để đưa ra tương tác phù hợp. Nhiều hãng máy ảnh hiện tại đã tích hợp tính năng nhận dạng khuôn mặt, nhận diện nụ cười để đưa ra bức ảnh đẹp nhất.

-Xác định độ tập trung: Trong các xe hơi hiện đại, hệ thống tự động của xe có thể phân tích đặc điểm khuôn mặt của người lái xe để đánh giá mức độ tập trung của họ. Đưa ra các cảnh báo phù hợp đảm bảo an toàn cho người lái xe.



Hình 1.1. Ứng dụng của nhận diện khuôn mặt

1.3 PHÁT HIỆN KHUÔN MẶT VÀ NHỮNG KHÓ KHĂN CỦA BÀI TOÁN NHẬN DẠNG KHUÔN MẶT

1.3.1 Phát hiện mặt người

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

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

1.3.2 Những khó khăn và thách thức của bài toán nhận dạng khuôn mặt

Bài toán nhận dạng mặt người là bài toán đã được nghiên cứu từ những năm 70. Tuy nhiên, đây là một bài toán khó nên những nghiên cứu hiện tại vẫn chưa đạt được kết quả mong muốn. Chính vì thế vấn đề này vẫn đang được nhiều nhóm trên thế giới quan tâm nghiên cứu. Khó khăn của bài toán nhận dạng mặt người có thể kể như sau:

- Tư thế, góc chụp: ảnh chụp khuôn mặt có thể thay đổi rất nhiều bởi vì góc chụp giữa camera và khuôn mặt. Chẳng hạn như: chụp thẳng, chụp xéo bên trái 45^0 , chụp từ trên xuống, chụp từ dưới lên,.... Với các tư thế khác nhau, các thành phần trên khuôn mặt như mắt, mũi, miệng có thể bị khuất một phần hoặc có thể khuất hết.

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

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

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

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

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

- Số lượng khuôn mặt trong ảnh, việc xác định khuôn mặt sẽ khó khăn hơn nhiều khi có sự xuất hiện đồng thời của nhiều khuôn mặt trong ảnh.

1.4 BÀI TOÁN PHÁT HIỆN MẶT NGƯỜI TRONG ẢNH

1.4.1 Giới thiệu tổng quát bài toán

Các bài toán phát hiện và nhận dạng khuôn mặt luôn gặp nhiều khó khăn, một trong các khó khăn chính đó là hướng nhìn (Face Pose) của khuôn mặt trong ảnh. Đa số các bài toán nhận dạng khuôn mặt chỉ có thể thực hiện được hoặc đạt kết quả chính xác nếu các bức ảnh chứa khuôn mặt ở tư thế thẳng đứng, hoặc có góc lệch nhỏ so với hướng chụp. Bài toán phát hiện hướng khuôn mặt người trong ảnh sẽ góp phần không nhỏ vào quá trình nhận dạng mặt người. Sử dụng kết quả của bài toán phát hiện hướng khuôn mặt người trong ảnh vào giai đoạn tiền xử lý hình ảnh, những bức ảnh chứa khuôn mặt người ở các tư thế khác nhau sẽ được qui chuẩn làm cơ sở đầu vào cho các hệ thống nhận dạng.

Không chỉ áp dụng cho các bước tiền xử lý hình ảnh trong nhận dạng khuôn mặt, phát hiện hướng mặt người trong ảnh còn được áp dụng vào các hệ thống tương tác người máy. Thông qua các biểu hiện khuôn mặt để điều khiển các chức năng của hệ thống. Ví dụ như thông qua hướng của khuôn mặt để điều khiển hướng của xe lăn trợ giúp người tàn tật hoặc ứng dụng trong các hệ thống cảnh báo độ tập trung của người lái xe.

Cùng trong một bức ảnh có thể có nhiều khuôn mặt ở những tư thế khác nhau. Sử dụng bài toán phát hiện hướng khuôn mặt nhằm xác định chính xác hướng của từng khuôn mặt trong ảnh. Qua đó kết hợp với quá trình tiền xử lý hình ảnh nhằm tạo ra những ảnh đầu vào tốt hơn rất nhiều cho các hệ thống nhận dạng.

Việc xác định tư thế khuôn mặt con người đặt ra nhiều vấn đề hơn so với các đối tượng khác bởi khuôn mặt người là một đối tượng động với nhiều hình thức thể hiện và màu sắc khác nhau.

Tương tác người máy (Human Computer Interaction - HCI) có thể được hỗ trợ rất tốt bằng việc sử dụng các yếu tố của việc nhận dạng cử chỉ, tư thế, cảm xúc tất cả những yêu cầu việc phát hiện theo dõi khuôn mặt.

Mặc dù có nhiều thuật toán khác nhau tồn tại thực hiện việc nhận diện khuôn mặt, mỗi thuật toán có điểm yếu và điểm mạnh riêng. Một số thuật toán sử dụng màu da, một số lại sử dụng đường nét và thậm chí phức tạp hơn liên quan đến mẫu, các mạng nơron hoặc các bộ lọc.... Những thuật toán này đều bị cùng một vấn đề là hiệu quả chưa cao. Một bức ảnh chỉ là một tập hợp của màu sắc và giá trị cường độ ánh sáng. Việc phân tích các điểm ảnh này để phát hiện khuôn mặt thì tốn thời gian để thực hiện và khó có thể hoàn thành bởi vì các biến thể về hình dạng, màu sắc trên khuôn mặt con người. Các điểm ảnh thường đòi hỏi phải được phân tích lại, được mở rộng quy mô và chính xác hơn. Theo đó Viola và Jones đã đề ra một thuật toán gọi là thuật toán nhận dạng Haar để nhanh chóng phát hiện bất kỳ đối tượng nào, bao gồm khuôn mặt con người sử dụng phương thức nhận dạng AdaBoost dựa trên các đặc tính Haar và không phải điểm ảnh.

1.4.2 Các hướng tiếp cận liên quan đến nhận dạng khuôn mặt

Có nhiều nghiên cứu tìm phương pháp xác định khuôn mặt người, từ ảnh xám đến ngày nay là ảnh màu. Dựa vào tính chất của các phương pháp xác định khuôn mặt người trên ảnh thì các phương pháp này được chia làm bốn hướng tiếp cận chính: dựa trên tri thức (Knowledge - Base), đặc trưng bất biến (Feature Invariant), đối sánh mẫu (Template

Matching) và dựa vào diện mạo (Appearance - Base).

- *Hướng tiếp cận dựa trên tri thức:* Trong hướng tiếp cận này, các phương pháp xác định mặt người được xây dựng dựa vào các luật, các luật này phụ thuộc vào tri thức của tác giả nghiên cứu về bài toán xác định khuôn mặt. Dễ dàng xây dựng các luật cơ bản để mô tả các đặc trưng của khuôn mặt và các quan hệ tương ứng. Ví dụ một khuôn mặt thường có hai mắt đối xứng qua trục thẳng đứng giữa khuôn mặt và có một mũi miệng. Các quan hệ đặc trưng có thể là quan hệ về vị trí và khoảng cách tương đối. Khó khăn của phương pháp này là làm sao chuyển tri thức của con người về khuôn mặt sang các luật cho máy tính một cách hiệu quả. Nếu các luật này quá tổng quát, đơn giản sẽ dẫn tới hiệu suất phát hiện khuôn mặt không cao, nhận diện nhầm nhưng các luật quá chi tiết thì có thể xác định thiếu các khuôn mặt trong ảnh. Ngoài ra rất khó để xác định được khuôn mặt với các tư thế khác nhau.

- *Hướng tiếp cận dựa trên các đặc trưng bất biến:* Đây là hướng tiếp cận theo kiểu bottom up. Dựa trên thực tế chúng ta có thể tìm ra những điểm không thay đổi của khuôn mặt khi ở các tư thế khác nhau, các điều kiện môi trường và ánh sáng khác nhau. Đã có nhiều nghiên cứu về hướng tiếp cận này, các đặc điểm không đổi trên khuôn mặt có thể kể tới như lông mày, mắt, mũi, miệng hay các đường viền trên tóc. Trên cơ sở này, ta xây dựng một mô hình thống kê mô tả quan hệ các đặc trưng và từ đó xác định sự xuất hiện của khuôn mặt trong bức ảnh. Khó khăn của phương pháp này là phải mô tả các đặc trưng không thay đổi và quan hệ giữa chúng phù hợp với điều kiện ánh sáng, độ nhiễu hoặc bị che khuất.

Có rất nhiều đặc trưng của khuôn mặt, có thể dựa vào các đặc trưng như hai mắt, hai lỗ mũi, miệng, phần nổi giữa mũi và miệng...

hoặc dựa vào đường viền của khuôn mặt.

Khuôn mặt người có những kết cấu riêng biệt mà có thể dùng phân loại so với các đối tượng khác. Từ kết cấu của khuôn mặt xây dựng lên một bộ phân loại, gọi là kết cấu giống của khuôn mặt (face-like texture).

Màu sắc da cũng là một tính chất được dùng để chọn các ứng viên có thể là khuôn mặt để thu hẹp dữ liệu giúp xác định khuôn mặt người.

Gần đây có nhiều nghiên cứu sử dụng các đặc trưng toàn cục như: màu sắc da người, kích thước và hình dáng để tìm các ứng viên khuôn mặt. rồi sau đó xác định khuôn mặt thông qua các đặc trưng cục bộ như lông mày, mắt, mũi, miệng...

- *Hướng tiếp cận dựa trên đối sánh mẫu*: Trong hướng tiếp cận này một mẫu khuôn mặt chuẩn được định nghĩa bằng tay trước hoặc được tham số hóa bằng một hàm số. Mẫu này được sử dụng để phát hiện khuôn mặt bằng cách quét nó qua ảnh và tính toán giá trị tương đồng cho mỗi vị trí. Việc xuất hiện một khuôn mặt tại một vị trí nào đó trong ảnh phụ thuộc vào giá trị tương đồng của điểm đó so với mẫu chuẩn. Thông qua các vị trí tương quan này mà hệ thống sẽ quyết định có hay không có tồn tại khuôn mặt trong ảnh. Hướng tiếp cận này có lợi thế là rất dễ cài đặt nhưng không hiệu quả khi tỷ lệ, tư thế và hình dáng thay đổi.

- *Hướng tiếp cận dựa trên diện mạo*: Ngược với phương pháp đối sánh mẫu sử dụng các mẫu được các chuyên gia định nghĩa trước, phương pháp này sử dụng các mẫu được rút trích qua một quá trình học. Nói cách khác, các thuật toán dựa trên máy học dùng các kỹ thuật phân tích thống kê và máy học để xấp xỉ một hàm phân lớp tuyến tính. Có rất nhiều mô hình máy học được áp dụng trong hướng tiếp cận này.

Trong hướng tiếp cận này, phương pháp xác xuất tiếp tục được sử dụng, một bức ảnh hoặc một vector (đặc trưng cho bức ảnh) được xem như một biến ngẫu nhiên. Biến ngẫu nhiên này sẽ được xác định là có mô tả các đặc tính của khuôn mặt hay không thông qua các xác suất có điều kiện $p(x/\text{face})$ và $p(x/\text{nonface})$. Có thể dùng các bộ phân loại Bayesian để phân loại các ứng viên là khuôn mặt hay không phải khuôn mặt. Tuy nhiên, việc cài đặt bộ phân loại Bayesian với chiều x lớn là một điều khá khó khăn, do đó có rất nhiều nghiên cứu theo hướng tiếp cận này quan tâm đến việc tham số hóa hay không tham số hóa các xác suất $p(x/\text{face})$ và $p(x/\text{nonface})$.

Một hướng tiếp cận khác là tìm một hàm biệt số, ví dụ như hàm mô tả mặt phẳng hay siêu phẳng quyết định (các ranh giới) để phân biệt hai lớp khuôn mặt và không phải khuôn mặt. Thông thường một bức ảnh được chiếu vào không gian có số chiều ít hơn sau đó dùng một hàm biệt số (dựa trên các độ đo khoảng cách) để phân loại hoặc xây dựng một mặt quyết định phi tuyến bằng mạng neural đa tầng. Ngoài ra có thể dùng SVM (Support Vector Machine) và các phương thức kernel chiếu các mẫu vào không gian có số chiều cao hơn để dữ liệu bị rời rạc hoàn toàn và từ đó có thể sử dụng mặt phẳng phân loại để phân loại các mẫu nào là khuôn mặt hay không phải là khuôn mặt.

1.4.3 Phát hiện mặt người trong ảnh dựa trên ngưỡng màu của da người và phương pháp trừ nền

1.4.3.1 Tách mặt người trong ảnh

Việc tách mặt người ra khỏi ảnh số là một bài toán gắn liền với hệ thống nhận dạng mặt người tự động.

Hiện nay có một số phương pháp để tách mặt người trong ảnh:

- Sử dụng ngưỡng màu đối với ảnh màu để xác định vùng có màu da của khuôn mặt người.
- Phương pháp trừ nền.

1.4.3.2 Sử dụng ngưỡng màu

Sử dụng ngưỡng màu là phương pháp thường được sử dụng với ảnh màu để tách vùng có màu da người. Để sử dụng phương pháp này, các hệ màu thường dùng là HSV, YcbCr và một số hệ màu khác. Ảnh đầu vào từ webcam hoặc camera thường ở hệ màu mặc định là RGB. Vì vậy cần phải sử dụng ma trận chuyển hệ màu để chuyển qua các hệ màu khác. Với hệ màu YcbCr ta sử dụng ma trận sau:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.797 & -74.203 & 112.000 \\ 112.000 & -93.786 & -18.214 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Thông thường với hệ màu YcbCr thì ngưỡng để xác định màu da người là:

$$77 \leq Cb \leq 127$$

$$133 \leq Cr \leq 173$$

Phương pháp nhận dạng khuôn mặt sử dụng ngưỡng màu da có một số ưu điểm như:

- Phương pháp đơn giản
- Có thể thực hiện trong ảnh có nhiều người, nhiều khuôn mặt.
- Có thể thực hiện với các ảnh có nền khác nhau.
- Có thể thực hiện với ảnh có mức sáng khác nhau.

Tuy nhiên phương pháp này cũng tồn tại nhiều nhược điểm như:

- Phải sử dụng ảnh màu, dẫn tới dung lượng lớn.
- Ngưỡng màu da là tương đối, vì làn da con người có nhiều màu khác nhau và màu sắc thể hiện trên ảnh số có thể bị thay đổi.

- Khi màu da trung màu nền sẽ sinh ra nhiễu. Vì sử dụng ảnh màu nên với những ảnh có kích thước lớn, việc tính toán cũng sẽ lớn và đòi hỏi tài nguyên phần cứng cao.

1.4.3.3 Phương pháp trừ nền

Về cơ bản, phương pháp này so sánh ảnh có khuôn mặt và ảnh không có khuôn mặt. Sự chênh lệch mức xám thể hiện sự xuất hiện của đối tượng.

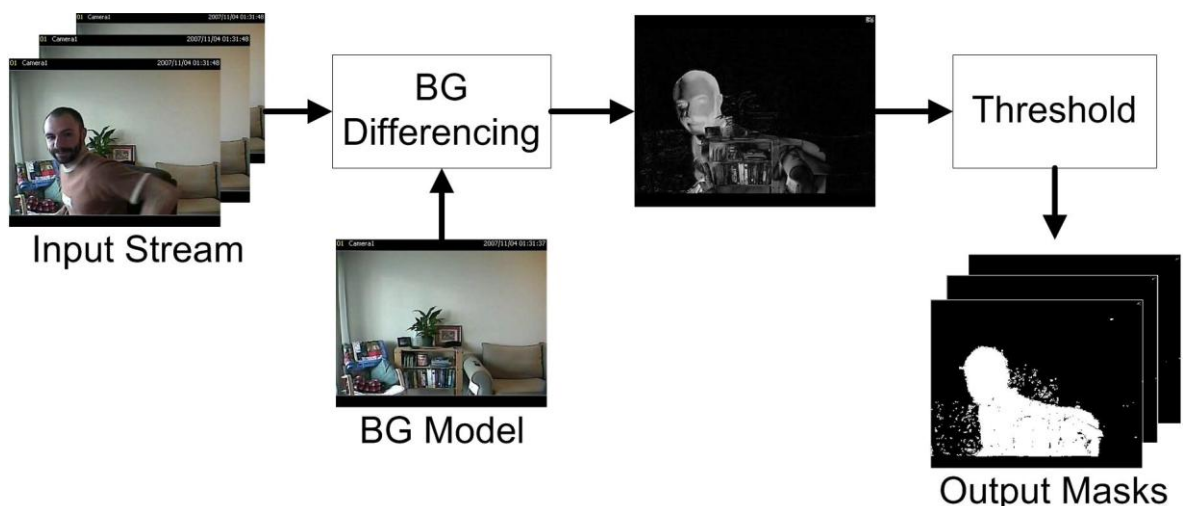
Trong phương pháp này, với hai điểm ảnh cùng một vị trí trên hai ảnh, nếu hiệu mức xám giữa chúng lớn hơn ngưỡng cho phép, ta xác nhận mức xám tại điểm đó có sự thay đổi và xác định được đối tượng.

Ưu điểm của phương pháp trừ nền là:

- Phương pháp đơn giản.
- Thời gian thực hiện ngắn
- Chỉ cần sử dụng ảnh xám.

Tuy nhiên nó cũng tồn tại một số nhược điểm:

- Chỉ ứng dụng với ảnh nền không đổi.
- Điều kiện ánh sáng không được thay đổi qua các lần chụp.



Hình 1.2. Phương pháp trừ nền với nền không đổi

1.4.4 Phát hiện mặt người trong ảnh dựa trên các đặc trưng Haar

Hai phương pháp được trình bày ở trên còn nhiều hạn chế, tỷ lệ phát hiện khuôn mặt trong ảnh chưa cao, tỷ lệ phát hiện nhầm lớn. Hiện nay, nhiều kỹ thuật trong tương tác người máy (HCI) yêu cầu những đặc trưng khởi tạo ban đầu. Với kỹ thuật nhận diện khuôn mặt cần chú ý nhiều đến các đặc điểm trên khuôn mặt.

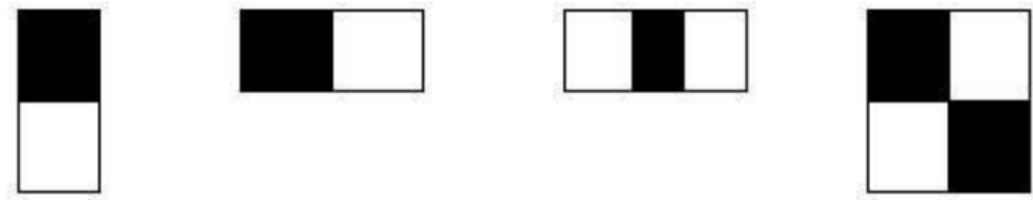
Mục tiêu của quá trình nhận dạng là nhận dạng dạng ảnh dựa trên các giá trị đặc trưng đơn giản. Hiện tại lớp nhận dạng Haar (Haar Cascade Classifier – HCC) ngày càng được chú ý với độ chính xác cao và hiệu quả thực thi tốt.

Haar Cascade Classifier kết hợp hiệu quả ba yếu tố. Đầu tiên nó làm việc trực tiếp với các giá trị xử lý ảnh, việc phát hiện của người sử dụng bao gồm một tập các đặc điểm đặc biệt có thể mang lại lợi ích cho việc tính toán sau này. Việc tiếp cận dựa trên các đặc điểm đặc biệt này sẽ giúp cho việc giảm các biến đổi trong một hàm và tăng sự biến đổi giữa các lớp với nhau. Thứ hai, nó sử dụng thuật toán Boost cho phép lựa chọn một lúc các đặc điểm vừa đủ cho việc đào tạo các nhận dạng này. Cuối cùng, việc tạo một cấu trúc lớp của kết quả nhận dạng ngày càng phức tạp hơn trong một chương trình phát hiện nhanh và hiệu quả.

1.4.4.1 Các đặc trưng Haar (Haar Like Feature)

Mục tiêu của quá trình nhận dạng dựa trên đặc trưng Haar là nhận dạng được ảnh dựa trên các giá trị đặc trưng đơn giản. Có nhiều lý do để đưa đến việc dùng các đặc trưng thay cho việc phân tích từng điểm ảnh mà lý do chính là các đặc trưng đó có thể được mã hóa thành dạng kiến thức học mà nếu thực hiện trên các điểm ảnh sẽ chậm hơn rất nhiều. Các hệ thống dựa trên các đặc tính sẽ nhanh hơn dựa trên các điểm ảnh.

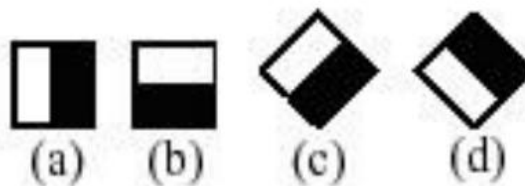
Viola và Jones đưa ra bốn đặc trưng cơ bản để xác định khuôn mặt người. Mỗi đặc trưng là sự kết hợp của hai hay ba hình chữ nhật “trắng” hoặc “đen như hình dưới:



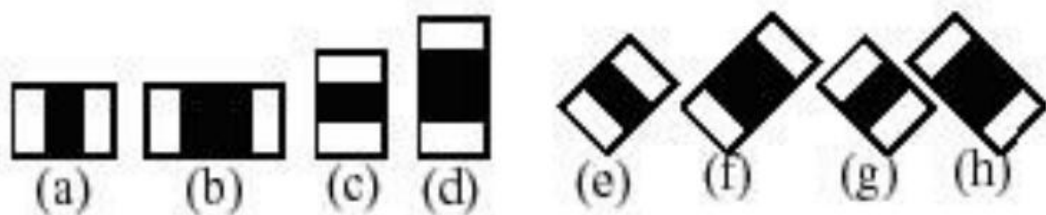
Hình 2.3: Bốn đặc trưng Haar-Like cơ bản

Để áp dụng các đặc trưng này vào bài toán xác định mặt người, bốn đặc trưng Haar-Like cơ bản được mở rộng và được chia thành ba tập đặc trưng sau:

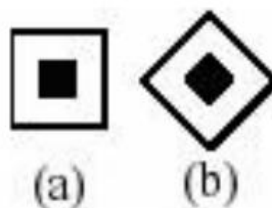
- Đặc trưng cạnh (Edge Features):



- Đặc trưng đường (Line Features):



- Đặc trưng xung quanh tâm (Center – Surround Features):

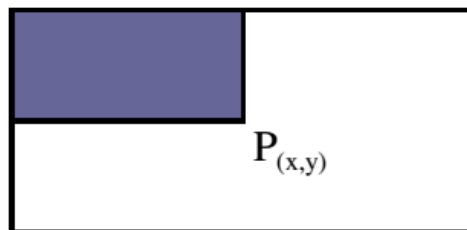


Hình 2.4: Các đặc trưng mở rộng của đặc trưng Haar-Like

Lợi ích của các đặc trưng Haar-Like là nó diễn đạt được tri thức về các đối tượng trong ảnh (bởi vì nó biểu diễn mối liên hệ giữa các bộ phận của đối tượng), điều mà bản thân từng điểm ảnh không thể diễn đạt được. Bản chất của việc tính các đặc trưng Haar-Like là tính sự chênh lệch giữa tổng các pixel vùng đen và tổng các pixel vùng trắng:

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

Như vậy có thể thấy rằng, để tính các giá trị đặc trưng Haar-Like, ta phải tính tổng của các vùng pixel trên ảnh. Nhưng để tính toán các giá trị đặc trưng Haar-Like cho tất cả các vị trí trên ảnh đòi hỏi chi phí tính toán rất lớn, không đáp ứng cho các ứng dụng đòi hỏi thời gian thực. Để khắc phục điều này, Viola và Jones đưa ra một khái niệm gọi là Integral Image để tính toán nhanh cho các feature cơ bản. Sau này được Lienhart kế thừa gọi là Integral Image là SAT – Summed Area Table và đưa ra thêm khái niệm RSAT – Rotated Summed Area Table dùng để tính toán nhanh cho các đặc trưng xoay một góc 45° . Integral Image là một mảng hai chiều có kích thước bằng kích thước của ảnh cần tính các đặc trưng Haar-Like, với mỗi phần tử của mảng này được tính bằng cách tính tổng các điểm ảnh phía trên (dòng - 1) và bên trái (cột - 1) của nó. Bắt đầu từ vị trí trên bên trái đến vị trí dưới, bên phải của ảnh, việc tính toán này chỉ đơn giản là phép cộng số nguyên cơ bản, do vậy tốc độ thực hiện rất nhanh.



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

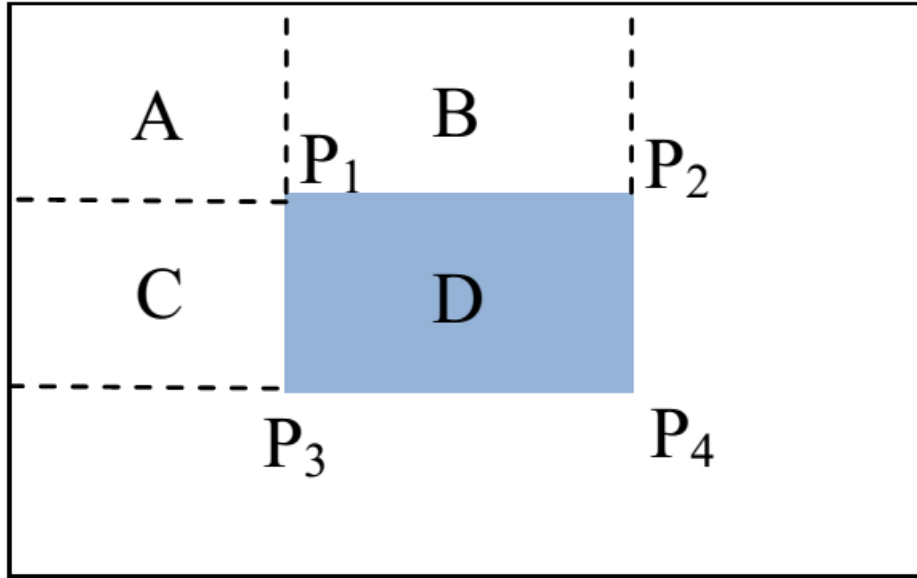
Hình 1.3. Cách tính Integral Image

Sau khi tính được Integral Image của ảnh, ta có thể tính tổng điểm ảnh của một vùng bất kì theo cách sau: (Giả sử ta cần tính tổng các điểm ảnh của vùng D như trong hình 2.6)

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

Với $A+B+C+D$ là các giá trị tại điểm P_4 trên Integral Image, tương tự vậy $A+B$ là giá trị tại điểm P_2 , $A+C$ là giá trị tại điểm P_3 và A là giá trị tại điểm P_1 . Như vậy ta có thể viết lại biểu thức tính D ở trên thành:

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



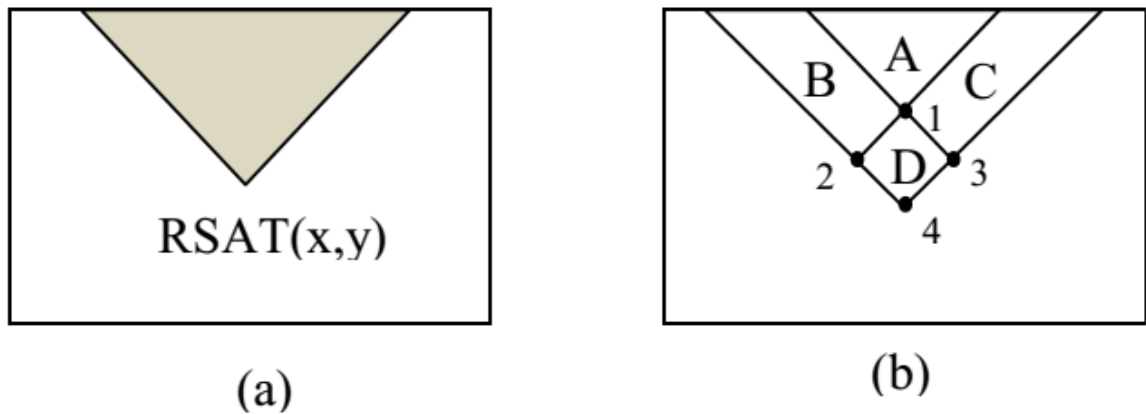
Hình 1.4. Cách tính nhanh tổng các điểm ảnh của vùng D trên ảnh

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

$$P(x,y) = \sum I(x',y')$$

Tổng pixel của một vùng bất kì trên ảnh vẫn được tính theo công thức:

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



Hình 1.5. Cách tính tổng điểm ảnh vùng D với đặc trưng xoay 45°

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

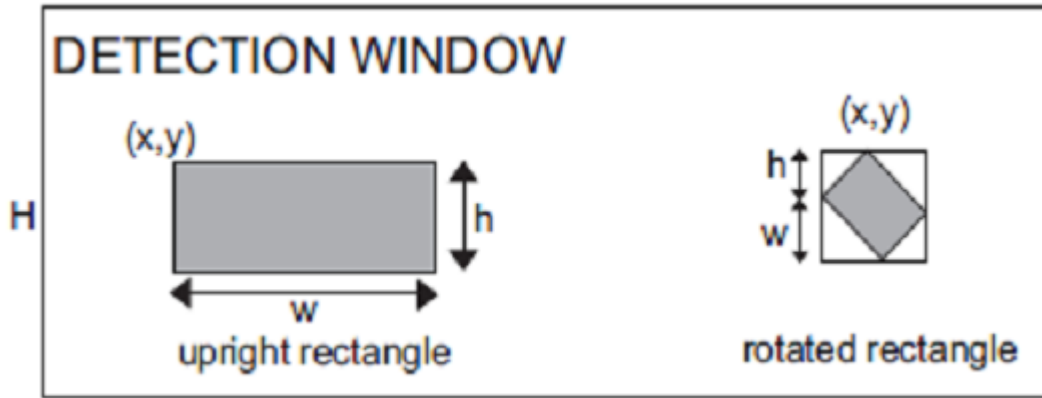
$$\text{Sum}(D) = 4 - 2 - 3 + 1$$

Ví dụ để phát hiện vị trí của mắt, lớp nhận dạng sẽ được huấn luyện dựa trên các đặc trưng cơ bản như các hình chữ nhật đen và trắng như hình dưới:



Hình 1.6. Sử dụng đặc trưng cơ bản để xác định vùng mắt

Với $\text{RecSum}(x, y, w, h, \Phi)$ là tổng các giá trị cường độ trên hình chữ nhật thẳng đứng hay xoay một góc Φ trong đó x, y, w, h, Φ lần lượt là tọa độ, các hướng và góc quay của hình chữ nhật.



Hình 1.7. Dựng thẳng, xoay các hình chữ nhật trong cửa sổ phát hiện

Đối với các hình chữ nhật theo phương thẳng đứng, $SAT(x,y)$ được sử dụng để tính toán đứng. Mỗi mục trong bảng được định nghĩa như là tổng cường độ ở các điểm ảnh trong hình chữ nhật thẳng đứng từ vị trí $(0,0)$ đến vị trí (x,y) theo công thức sau:

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

Sau khi đã tính toán trên các điểm ảnh của hình chữ nhật, SAT cho phép tính toán tổng điểm ảnh trên hình chữ nhật thẳng đứng với công thức:

$$RecSum(x,y,w,h,\Phi=0) = SAT(x-1,y-1) + SAT(x+w-1,y+h-1) - SAT(x+w-1,y-1) - SAT(x-1,y+h-1)$$

Đối với hình chữ nhật quay một góc Φ , ta sử dụng giá trị $RSAT(x,y)$ tại một điểm ảnh trong hình chữ nhật để tính toán $RecSum(x,y,w,h,\Phi)$

$$RSAT(x,y) = \sum_{|x-x'| \leq y-y', y' \leq y} i(x', y')$$

Việc tính toán các điểm ảnh trong hình chữ nhật quay góc Φ sẽ được tính theo biểu thức:

$$RecSum(x,y,w,h,\Phi) = RSAT(x-h+w,y+w+h-\tan\Phi) + RSAT(x,y-\tan\Phi) - RSAT(h-x,y+h-\tan\Phi) - RSAT(x+w-1,y+w-\tan\Phi)$$

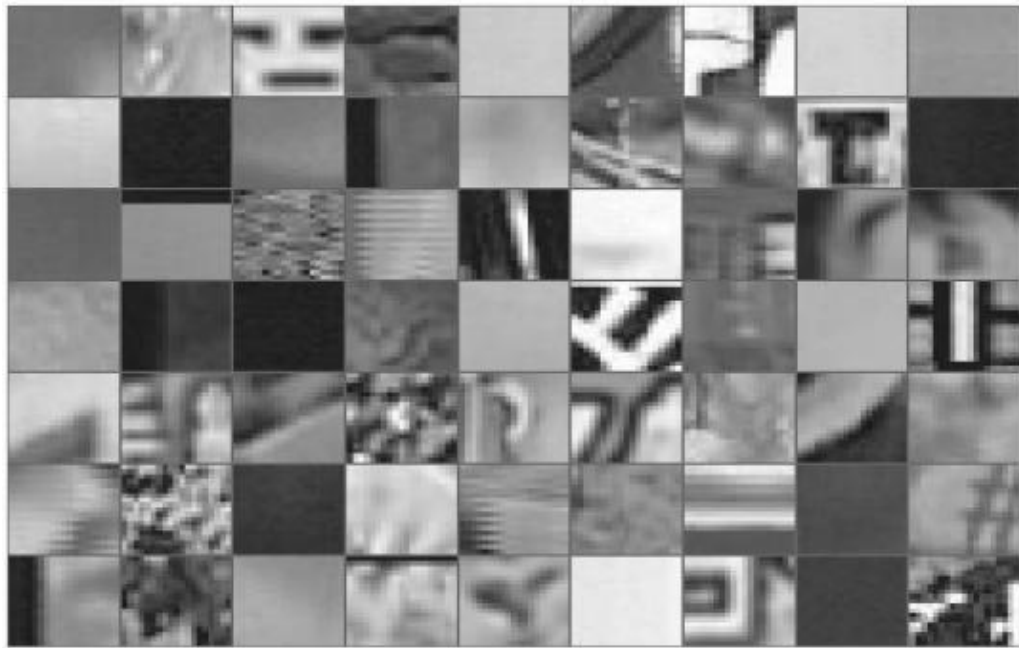
1.4.4.2 Huấn luyện các bộ nhận dạng Haar

Việc xác định các đặc điểm trên khuôn mặt như mắt, mũi, miệng yêu cầu các bộ nhận dạng Haar phải được huấn luyện. Trong quá trình huấn luyện các nhận dạng, các thuật toán AdaBoost và thuật toán Haar-Feature sẽ được thực hiện. Intel đã phát triển một thư viện mã nguồn mở giúp việc này được thực hiện dễ dàng trên máy tính với thư viện Open Computer Vision Library (OpenCV). OpenCV được dùng để kết nối các ứng dụng và được sử dụng trong một số lĩnh vực như tương tác người máy, tự động, sinh trắc học, xử lý ảnh và một số lĩnh vực khác.

Để huấn luyện các bộ nhận dạng này, tập hợp các ảnh ban đầu sẽ được nhận dạng. Tập ảnh đó bao gồm các ảnh chứa hoặc không chứa đối tượng cần phát hiện, ở đây là các đặc điểm trên khuôn mặt như mắt, mũi, miệng sẽ được xác định. Tập hợp các ảnh này bao gồm các mẫu âm (Negative Sample) không chứa đối tượng cần xác định và bao gồm các mẫu dương (Positive Sample) bao gồm các thể hiện của đối tượng cần xác định.



Hình 1.8. Một số mẫu dương (Positive Sample)
dùng cho việc xác định khuôn mặt



Hình 1.9. Một số mẫu âm (Negative Sample)

dùng cho việc xác định khuôn mặt

Vị trí của đối tượng trong một mẫu dương được xác định bởi: tên ảnh, điểm ảnh phía trên bên trái, , chiều cao, chiều rộng của đối tượng. Thư viện sẵn có của OpenCV sử dụng ít nhất 5000 mẫu âm để huấn luyện.

Để tạo ra các đặc điểm thiết thực nhất cho việc xác định các đặc điểm trên khuôn mặt, tập các mẫu dương ban đầu phải đại diện cho sự khác nhau bao gồm: chủng tộc, giới tính, tuổi tác. Một trong những nguồn mẫu dương tốt trong thực tế là cơ sở dữ liệu về công nghệ nhận dạng khuôn mặt của viện tiêu chuẩn công nghệ (NIST). Cơ sở này chứa đựng khoảng 10.000 ảnh với khoảng 1000 ảnh người dưới các điều kiện ánh sáng, tư thế, góc quay khác nhau. Ba lớp nhận dạng được đào tạo chỉ mặt, mũi và miệng. Sau khi các lớp nhận dạng này đã được đào tạo và áp dụng và các tập ảnh khác nhau trong cơ sở dữ liệu FERET ta sẽ được tỉ lệ các định các đặc điểm đó như bảng dưới đây.

Facial Feature	Positive Hit Rate	Negative Hit Rate
Eyes	93%	23%
Nose	100%	29%
Mouth	67%	28%

Bảng 1.1. Độ chính xác của bộ nhận dạng

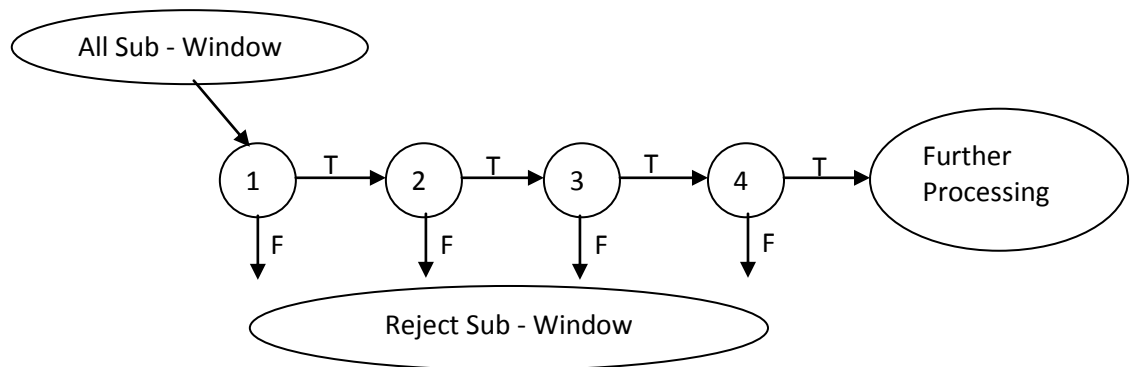
1.4.4.3 Chuỗi các bộ nhận dạng Haar

Quá trình huấn luyện một chuỗi phát hiện liên quan đến việc kết hợp của hai vấn đề. Chuỗi phát hiện với nhiều đặc trưng hơn sẽ nhận dạng chính xác hơn, tỉ lệ nhận dạng nhầm lẫn mẫu âm sẽ ít hơn. Tuy nhiên việc tăng số đặc trưng sẽ làm tốc độ tính toán chậm hơn, đòi hỏi thời gian tính toán phải nhiều hơn. Vì vậy ta phải đưa ra một định nghĩa cấu trúc tối ưu mà trong đó số tầng, số đặc trưng mỗi tầng và ngưỡng của mỗi tầng phải được dung hòa để tối thiểu hóa số đặc trưng. Việc tìm ra cấu trúc tối ưu này là rất khó khăn.

Trên thực tế, người ta xây dựng một cấu trúc đơn giản để tạo ra một nhận dạng hiệu suất cao với mục tiêu là giảm tối thiểu việc nhận dạng nhầm là dương và tối đa hiệu suất phát hiện đúng. Mỗi tầng được thêm các đặc trưng cho đến khi tỉ lệ nhận dạng nhầm là dương và tỉ lệ tìm kiếm đúng đạt được, được kiểm tra bằng tập mẫu kiểm chứng.

Trong hầu hết các trường hợp các định các đối tượng, chúng ta chỉ sử dụng một phần nhỏ của bức ảnh. Theo đó để loại bỏ các vùng không phải là đối tượng được các định nhanh chóng và chỉ tập trung vào những vùng liên quan, để kiểm tra tất cả các cửa sổ một cách triệt để. Cấu trúc lớp cho phép tiếp cận như vậy. Toàn bộ quá trình phát hiện có thể mô tả bằng một cây quyết định suy giảm dần. Nó bao gồm N giai đoạn trong một loạt các lớp nhận dạng được kết nối dùng để phân biệt các đối tượng được xác định và nền ảnh. Mỗi kết quả dương của bộ nhận dạng trước sẽ

kích hoạt việc đánh giá các nhận dạng sau và kết quả dương của bộ nhận dạng này lại kích hoạt cho bộ nhận dạng sau nữa. Mỗi giai đoạn được huấn luyện để đạt được tỷ lệ TP (True Positive) gần bằng 1 và tỉ lệ FP (False Positive) gần bằng 0,5. Các cửa sổ nhận dạng xác thực được thông qua các gian đoạn kế tiếp nhau. Bất cứ kết quả âm của tầng nhận dạng nào cũng sẽ loại bỏ khung hình đó.



Hình 1.10. Chuỗi các nhận dạng theo tầng

Trong 1 ảnh đơn bất kì, một lượng lớn các khung hình phụ sẽ có kết quả là âm. Do đó chuỗi phát hiện cố gắng loại bỏ càng nhiều khung hình âm trong những tầng đầu. Các bộ phận sau sẽ được huấn luyện trên các mẫu đã được thông qua ở tầng trước.

Điều này dẫn tới công việc của tầng sau sẽ khó hơn ở tầng trước do các mẫu ở tầng sau thường khó nhận dạng hơn, làm tỷ lệ nhận dạng nhầm ở các tầng này cao hơn so với tầng ban đầu.

1.4.4.4 Thuật toán ADABOOST

AdaBoost (Adaptive Boost) là một bộ phân loại phi tuyến mạnh cải tiến từ thuật toán Boosting, giúp đẩy nhanh việc tạo ra bộ phân loại mạnh (strong classifier) bằng cách chọn các đặc trưng tốt Haar-like trong bộ phân loại yếu (weak classifier) và kết hợp chúng lại tuyến tính để

hình thành 1 bộ phân loại mạnh (strong classifier) bằng cách sử dụng các trọng số (weight) để đánh dấu các mẫu khó nhận dạng.

Cho trước một vài thuật toán học yếu, người ta áp dụng kỹ thuật tăng cường để tạo ra thuật toán học mạnh hơn. AdaBoost là thuật toán cho phép ta có thể làm được điều trên. AdaBoost gồm có hai giai đoạn chính như sau:

Giai đoạn 1: Tạo ra các giả định yếu (weak hypotheses) từ các thuật toán học yếu.

Giai đoạn 2: Tạo ra các thuật toán học mạnh từ các giả định yếu.

Với một tập dữ liệu huấn luyện, một giả định yếu được khởi tạo như sau:

Đầu vào (input): Cho một tập dữ liệu huấn luyện, N cặp (x_i, y_i) , x_i là các vector đặc trưng Haar, y_i là giá trị đầu ra mong muốn luôn luôn mang giá trị bằng $+1$ hoặc -1 (trong đó $+1$ là đối tượng (object), -1 không phải là đối tượng (background)), và số lượng vòng lặp là T .

Đầu ra(output) tồn tại một hàm $f_T(x)$ có thể được sử dụng để phân lớp các đặc trưng của Vector x .

Nếu $f_T(x) < 0$ thì x được phân lớp là -1 .

Nếu $f_T(x) > 0$ thì x được phân lớp là $+1$.

Khởi tạo (Initialization): Cho trọng số $W_i = \frac{1}{N}$

Lặp (Iterate): For $t=1, 2, \dots, T$ tính giả định (hypothesis) h_t , trọng số tốt (goodness) α_t , và đồng thời cập nhật lại các trọng số W_1, \dots, W_N theo các bước sau:

Bước 1: Chọn ngẫu nhiên một tập con S_t trong tập dữ liệu huấn luyện. Trong trường hợp này chính là trọng số W_i .

Bước 2: Tính toán giả định h_t bằng cách sử dụng bộ phân lớp yếu cho S_t .

Bước 3: Tính toán sai số của trọng số huấn luyện ϵ_t của h_t .

$$\epsilon_t = \sum_{y_i \neq h_t(x_i)} P_i$$

Nếu $\epsilon_t \geq 0.5$ thì quay lại bước lặp.

Nếu $\epsilon_t = 0$ thì đây không phải là một phân lớp yếu, thực hiện tăng thêm số lượng mẫu huấn luyện.

Bước 4: Tính toán trọng số tốt α_t của h_t :

$$\alpha_t = \frac{1}{2} \ln \frac{1-\epsilon_t}{\epsilon_t}$$

Bước 5: Cập nhật lại trọng số

$$\text{Với } q_i = \begin{cases} e^{-\epsilon_t} & \text{nếu } h_t(x_i) = y_i \\ e^{\epsilon_t} & \text{nếu } h_t(x_i) \neq y_i \end{cases}$$

$$\text{Và } q_i \text{ mới: } q_i = \frac{W_i q_i}{Z_t} = \frac{W_i e^{\alpha_t y_i h_t(x_i)}}{Z_t}$$

Trong đó Z_t là thừa số chuẩn hóa sao cho $\sum_i W_i = 1$

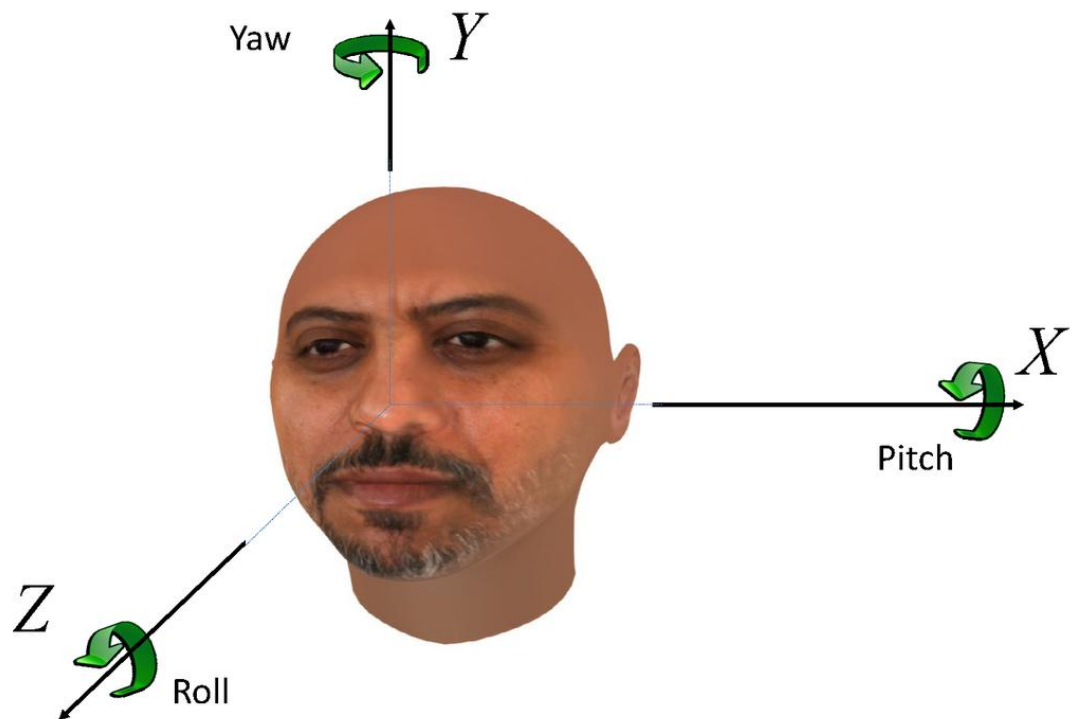
Kết thúc (Termination):

$$f_T(x) = \sum_{t=1}^T \alpha_t h_t(x)$$

Tính chất quan trọng nhất của AdaBoost là có f_{t+1} tốt hơn f_t .

1.4.5 Mô hình 3D của đầu người

Các nghiên cứu trước đây về nhận diện khuôn mặt tập trung chủ yếu vào việc xác định khuôn mặt trong không gian hai chiều, dựa trên các tập ảnh hoặc video. Tuy nhiên trên thực tế thì mô hình đầu người chuyển động trong một không gian ba chiều. Trong báo cáo tác giả đưa ra mô hình đầu người trong không gian ba chiều với ba góc quay Euler xung quanh ba trục trục giao.



Hình 1.11. Mô hình 3D của đầu người

CHƯƠNG 2

CÁC PHƯƠNG PHÁP XÁC ĐỊNH HƯỚNG MẶT NGƯỜI TRONG ẢNH

2.1 PHƯƠNG PHÁP XÁC ĐỊNH HƯỚNG MẶT NGƯỜI TRONG ẢNH

2.1.1 Xác định hướng mặt người dựa trên tâm của hai mắt

2.1.1.1 Xác định tâm mắt

Để thực hiện nhiệm vụ tìm khuôn mặt và góc quay của khuôn mặt, ta tiến hành tìm tâm hai mắt để xác định vị trí khuôn mặt, từ đó suy ra góc lệch của khuôn mặt so với tư thế chuẩn.

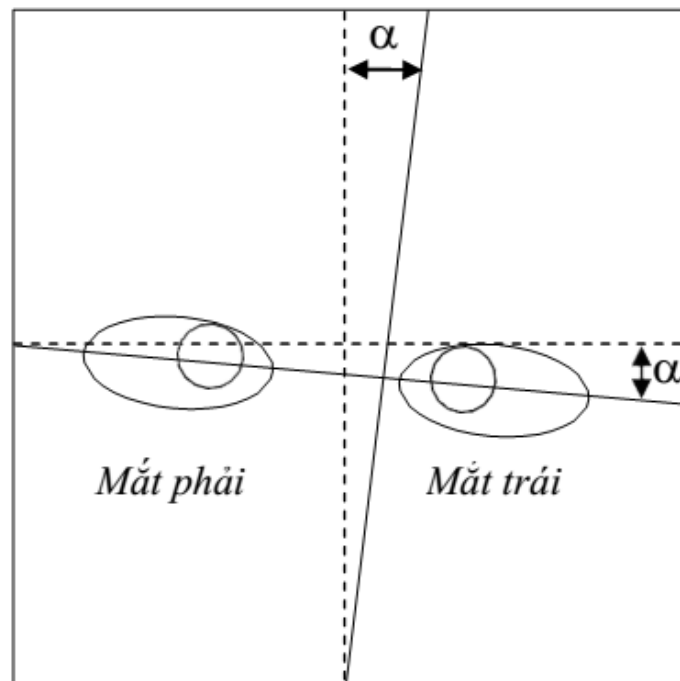
Để xác định vùng mắt, ta sử dụng các đặc trưng Haar duyệt nhanh toàn bộ ảnh để nhận diện khuôn mặt, sau đó duyệt riêng từng khuôn mặt vừa tìm được để xác định mắt. Dựa vào một số đặc điểm của mắt như khoảng cách giữa hai mắt, độ chênh chiều cao của hai mắt, v...v... ta đưa ra các phương thức cho phép xác định chính xác vị trí của mắt trong khuôn hình.

Sau khi xác định được vị trí của mắt, ta sử dụng phương pháp tìm tâm của một vùng để xác định tâm mắt. Tâm mắt sẽ là tâm của hình chữ nhật bao quanh mắt.

Từ vị trí tâm hai mắt, thực hiện phép xoay ảnh, ta xác định được góc nghiêng của khuôn mặt so với phương thẳng đứng.

2.1.1.2 Xoay ảnh và xác định góc nghiêng

Từ tọa độ tâm hai mắt, ta xác định được góc nghiêng của khuôn mặt so với phương thẳng đứng.



Hình 2.1. Xác định góc nghiêng dựa trên tâm hai mắt

Giả sử tâm của mắt trái có tọa độ (x_1, y_1) và mắt phải có tọa độ (x_2, y_2) . Với α là góc quay của khuôn mặt so với phương thẳng đứng.

$$\operatorname{tg} \alpha = \frac{y_2 - y_1}{x_2 - x_1}$$

Phương pháp để tính góc nghiêng của khuôn mặt dựa trên vị trí tâm của hai mắt là tương đối đơn giản, tuy nhiên phương pháp này chỉ xác định được góc nghiêng của khuôn mặt so với phương thẳng đứng. Với một số tư thế đặc biệt như xoay qua trái, xoay phải thì phương pháp này bộc lộ nhiều hạn chế.

2.1.2 Xác định hướng mặt người dựa trên các đặc tính về mắt, mũi, miệng





2.1.2.1 Xác định các đặc tính mắt, mũi, miệng

Phương pháp xác định mặt người trong ảnh đã được trình bày ở phần trên. Tương tự vậy, để xác định vị trí mũi, miệng, trước tiên ta duyệt toàn bộ ảnh để xác định khuôn mặt sau đó sử dụng các đặc trưng Haar khuôn hình chứa khuôn mặt để xác định vị trí mũi, miệng.

Do đặc tính đặc biệt về hình dạng nên việc xác định mũi trong khuôn mặt có kết quả chính xác rất cao, sử dụng một số đặc điểm của mũi như vị trí của mũi nằm bên dưới hai mắt, nằm giữa khuôn mặt để nâng cao hiệu suất tìm kiếm.

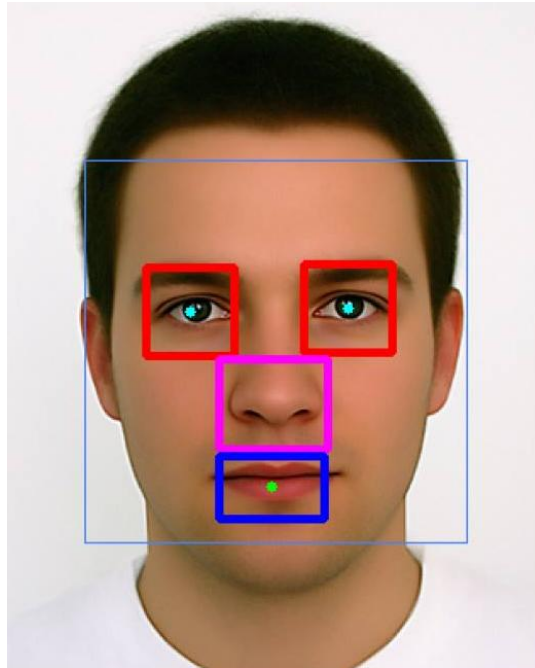
Việc xác định vị trí của miệng cũng tương đối đơn giản, dù miệng mang một số đặc tính về hình học khá giống với mắt nhưng ta cũng sẽ sử dụng vị trí tương quan của nó so với mũi để nâng cao độ chính xác.

Trong báo cáo, các lớp nhận dạng được xây dựng trong thư viện OpenCV được sử dụng để xác định bốn nhận dạng gồm khuôn mặt, mắt, mũi, miệng với bốn file xml.

STT	Xác định vị trí	Tên file
1	Khuôn mặt	 "haarcascade_frontalface_alt.xml"
2	Mắt	 "haarcascade_eye.xml"
3	Mũi	 "haarcascade_mcs_nose.xml"
4	Miệng	 "mouth.xml"

Bảng 2.1. Bảng liệt kê các file xml

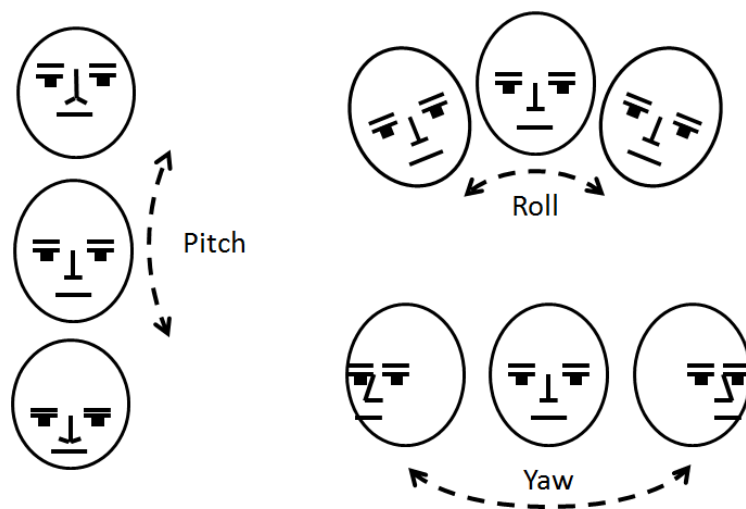
Kết quả của việc nhận dạng ta sẽ vẽ được các hình chữ nhật với các màu khác nhau bao quanh các điểm cần xác định



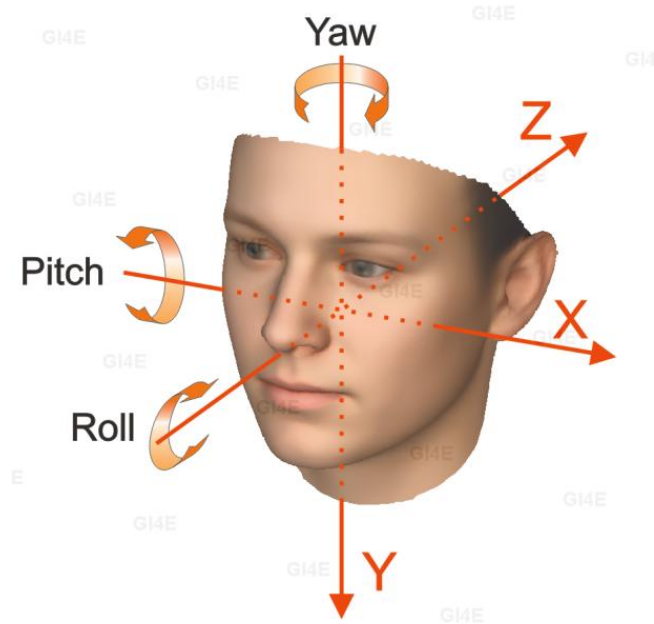
Hình 2.2. Kết quả sau khi phát hiện các đặc điểm

2.1.2.2 Mô hình mặt người trong không gian 3D.

Chúng ta biết rằng đầu người được xem là một khối 3D trong hệ tọa độ 3D. Do đó các biến của tư thế đầu cũng thỏa mãn các mẫu chung của các chuyển động cứng nhắc. Các ảnh khuôn mặt theo các tư thế khác nhau có thể xem như một phép chiếu 2D cho các góc quay khác nhau quanh trung tâm đầu.



Hình 2.3. Các góc quay của khuôn mặt



Hình 2.4. Hệ tọa độ 3D của khuôn mặt

2.1.2.3 Tính toán góc quay 3D của khuôn mặt

Gọi góc quay theo phương Yaw (Hình 2.3) là α , góc quay theo phương Pitch (hình 2.3) là β và góc quay theo phương Roll (hình 2.3) là γ .

Xác định góc quay theo hướng α

Góc quay $\alpha = 0$ khi nhìn thẳng và thay đổi trong khoảng từ $[-\frac{\pi}{2}, \frac{\pi}{2}]$ khi đầu người quay từ trái sang phải.

Khoảng cách từ biên trái của mặt đến tâm của miệng (Tâm của miệng được tính bằng cách lấy tâm của hình chữ nhật bao quanh miệng) có thể được tính theo công thức:

$$D_{ML} = |P_M(x) - P_{FL}(x)| \quad (2.1)$$

Với P_M là vị trí theo chiều ngang của môi và P_{FL} là vị trí của biên trái của khuôn mặt (được xác định là cạnh biên trái của hình chữ nhật bao quanh mặt).

Khoảng cách từ biên phải của mặt đến tâm của miệng có thể được tính theo công thức:

$$D_{MR} = |P_M(x) - P_{FR}(x)| \quad (2.2)$$

Với P_{FR} là biên phải của khuôn mặt (được xác định là cạnh biên phải của hình chữ nhật bao quanh mặt).

Bây giờ chúng ta có thể tính được hiệu của hai khoảng cách này:

$$D_{HS} = D_{ML} - D_{MR} \quad (2.3)$$

Như vậy một phương trình tuyến tính có thể được thiết lập để mô tả mối quan hệ giữa α và D_{HS} như sau:

$$\frac{D_{HS}}{D_{LR}} = k \frac{\alpha}{\theta} + b \quad (2.4)$$

Với k , θ và b là hằng số. Chúng ta có thể có được các giá trị này bằng cách thử nghiệm ở hai trường hợp đặc biệt. Trường hợp nhìn thẳng ($\alpha=0$) và trường hợp quay phải 90° ($\alpha=\pi/2$).

Trong trường hợp nhìn thẳng: $\alpha=0$, $P_M(x)=0$, $D_{MR}=\frac{D_{LR}}{2}$ và $D_{LR}=\frac{D_{LR}}{2}$

Trong trường hợp quay phải 90° : $\alpha=\pi/2$, $D_{ML}=D_{LR}$, $D_{MR}=0$

Từ đây ta có thể có các nhóm phương trình sau:

$$\begin{aligned} \left(\frac{\frac{D_{LR}-D_{LR}}{2}}{D_{LR}} - b \right) \cdot \theta &= 0 \\ \left(\frac{\frac{D_{LR}-0}{2}}{D_{LR}} - b \right) \cdot \theta &= \pi/2 \end{aligned} \quad (2.5)$$

Với k , θ và b có thể được tính từ các nhóm: $b=0$, $\theta = k \cdot (\pi/2)$. Từ đó tính được góc quay α theo trục Y là:

$$\alpha = \frac{\pi}{2} \cdot \frac{2P_M(x) - D_{LR}}{D_{LR}} \quad (2.6)$$

Xác định góc quay β theo hướng Pitch

Tương tự như xác định góc quay α , việc xác định góc quay β cũng được thực hiện bằng cách so sánh vị trí tương quan của miệng với cạnh

trên và cạnh dưới hình chữ nhật bao quanh mắt. Góc $\beta=0$ khi nhìn thẳng và β nằm trong khoảng $[-\pi/2, \pi/2]$ khi đầu ngẩng lên hoặc cúi xuống.

Khoảng cách giữa miệng và ranh giới trên của khuôn mặt có thể được tính bằng phương trình:

$$D_{MT} = |P_M(y) - P_{FT}(y)| \quad (2.7)$$

Với k , θ và b là hằng số.

Thông thường ở vị trí thẳng đứng, miệng con người không phải ở trung tâm khuôn mặt vì vậy đầu tiên ta phải chuyển đổi tọa độ của miệng lên tâm điểm.

$$P'_M(y) = P_M(y) - D_{MC}$$

Tương tự ta sẽ có phương trình:

$$\beta = \frac{\pi}{2} \cdot \frac{D_{VS}}{D_{TB}} \quad (2.8)$$

$$\text{Với } D_{VS} = D_{MT} - D_{MB} = 2P_M(y) - D_{TB} - 2D_{MC} \quad (2.9)$$

Như vậy hướng quay của khuôn mặt theo trục Y sẽ được tính:

$$\beta = \frac{\pi}{2} \cdot \frac{2P_M(y) - D_{TB} - 2D_{MC}}{D_{TB}} \quad (2.10)$$

Trong phương trình (2.6) D_{LR} là chiều rộng của khuôn mặt, ở phương trình (2.10) thì D_{TB} là chiều cao của khuôn mặt, $P_M(x,y)$ là vị trí của miệng và có thể được xác định dễ dàng bằng các thuật toán phát hiện khuôn mặt. D_{MC} là một hằng số và được tính bằng thực nghiệm. Sử dụng dữ liệu khuôn mặt có thể ước lượng được $D_{TB} = 7/40$.

Xác định góc quay γ (Roll)

Như đã trình bày ở phần trên, khi mặt quay về trục z thì sẽ tồn tại một khoảng cách độ lệch giữa hai mắt theo trục thẳng đứng. Giả định rằng khi mặt người nhìn thẳng giá trị độ lệch đó bằng 0 và khi đầu quay theo trục Z thì giá trị độ lệch đó có thể được tính theo phương trình:

$$\gamma = \arctan\left(\frac{D_{LR}(x)}{D_{LR}(y)}\right) \quad (2.11)$$

Với $D_{LR}(x)$ và $D_{LR}(y)$ lần lượt là độ lệch ngang và dọc giữa hai mắt. Chúng có thể được biểu diễn như sau:

$$D_{LR}(x) = P_{LE}(x) - P_{RE}(x) \quad (2.12)$$

$$D_{LR}(y) = P_{LE}(y) - P_{RE}(y) \quad (2.13)$$

Với $P_{LE}(x,y)$ là tọa độ của tâm mắt trái và $P_{RE}(x,y)$ là tọa độ của tâm mắt phải.

$$\gamma = \frac{P_{LE}(x) - P_{RE}(x)}{P_{LE}(y) - P_{RE}(y)}$$

2.2 XÁC ĐỊNH GÓC NHÌN CỦA KHUÔN MẶT DỰA TRÊN CÁC ĐIỂM MỐC TRÊN KHUÔN MẶT.

2.2.1 Phát hiện các điểm mốc (Land marks) trên khuôn mặt.

Phát hiện các điểm mốc trên khuôn mặt là việc định vị các điểm chuẩn trên hình ảnh khuôn mặt. Những đặc điểm này và hình dạng khuôn mặt sẽ giúp ích rất nhiều cho việc phân tích khuôn mặt trong đó có việc xác định góc quay của khuôn mặt trong ảnh.

Gần đây đã có nhiều thuật toán, thư viện hỗ trợ cho việc xác định các điểm mốc, xác định biên của khuôn mặt. Tuy nhiên vẫn còn nhiều thách thức nếu góc quay lớn trên 60^0 hoặc điều kiện ánh sáng yếu, v...v...

Tuy nhiên nếu ta coi các điểm mốc trên những phần tự che khuất của khuôn mặt như những điểm bị che khuất, chúng ta có thể xác định những mặt người với góc quay lớn như là một trường hợp đặc biệt của ảnh bị che khuất và xử lý nó như những trường hợp bình thường.

Trong nghiên cứu, tác giả dựa trên kết quả bài báo “*One Millisecond Face Alignment with an Ensemble of Regression Trees*” của nhóm tác giả *Vahid Kazemi and Josephine Sullivan - KTH, Royal Institute of Technology - Computer Vision and Active Perception Lab -*

Teknikringen 14, Stockholm, Sweden đăng trong tạp chí CVPR2014 để xác định các điểm mốc trên khuôn mặt.

Bài báo trên trình bày một thuật toán để ước lượng chính xác vị trí các điểm mốc trên khuôn mặt một cách hiệu quả. Tương tự các phương pháp trước, phương pháp sử dụng một chuỗi các biến hồi quy.

- Phân tầng hồi quy:

Cho $x_i \in \mathbb{R}^2$ với tọa độ (x, y) là điểm mốc thứ i trong ảnh I .

Vector $S = (x_1^T, x_2^T, \dots, x_p^T)^T \in \mathbb{R}^{2p}$ biểu thị tọa độ của tất cả các điểm mốc P trên khuôn mặt. \hat{S}^t là biểu thị cho ước lượng hiện tại của S .

$$\hat{S}^{(t+1)} = \hat{S}^{(t)} + r_t(I, \hat{S}^{(t)})$$

Các điểm quan trọng của phân tầng là r_t dự đoán dựa trên các đặc trưng như giá trị cường độ pixel.

- Học mỗi ước lượng hồi quy của phân tầng:

Giả sử có dữ liệu huấn luyện $(I_1, S_1) \dots (I_n, S_n)$ với mỗi I_i là khuôn mặt và S_i là vector khuôn dạng. Để học hàm ước lượng hồi quy đầu tiên r_0 của phân tầng chúng ta tạo ra 3 dữ liệu huấn luyện về hình ảnh khuôn mặt, ước lượng hình dạng ban đầu và bước cập nhật mục tiêu, đó là $I_{\pi i}$, $\hat{S}_i^{(0)}$, $\Delta S_i^{(0)}$ với:

$$I_{\pi i} \in \{1, \dots, n\}$$

$$\hat{S}_i^{(0)} \in \{S_1, \dots, S_n\} \setminus S_{\pi i}$$

$$\Delta S_i^{(0)} = S_{\pi i} - \hat{S}_i^{(0)}$$

$$\hat{S}_i^{(t+1)} = \hat{S}_i^{(t)} + r_t(I_{\pi i}, \hat{S}_i^{(t)})$$

$$\Delta S_i^{(t+1)} = S_{\pi i} - \hat{S}_i^{(t+1)}$$

Từ đó đưa ra thuật toán học r_t trong phân lớp

- Bước 1: khởi tạo

$$f_0(I, \hat{S}^{(t)}) = \arg \min_{\gamma \in \mathbb{R}^{2p}} \sum_{i=1}^N \|\Delta S_i^{(t)} - \gamma\|^2$$

- Bước 2: Cho k chạy từ 1, ..., K

Cho i chạy từ 1, ..., N

$$r_{i,k} = \Delta S_i^{(t)} - f_{k-1}(I_{\pi i}, \hat{S}_i^{(t)})$$

$$\text{cập nhật: } f_k(I, \hat{S}_i^{(t)}) = f_{k-1}(I, \hat{S}_i^{(t)}) + v g_k(I, \hat{S}_i^{(t)})$$

$$\text{Đầu ra } r_t(I, \hat{S}_i^{(t)}) = f_k(I, \hat{S}_i^{(t)})$$

- Cây hồi quy cơ sở:

Cốt lõi của hàm hồi quy r_t là cây hồi quy cơ sở phù hợp các mục tiêu trong thuật toán gradient boosting.

2.2.2 Chuyển đổi ma trận xoay qua 3 góc Euler

Như đã trình bày ở trên, hiện nay có nhiều thuật toán cho phép xác định các điểm mốc, xác định biên trên khuôn mặt. Xác định này sẽ được đặc trưng bởi một ma trận xoay 3x3. Để xác định góc xoay của khuôn mặt từ ma trận này ta phải thực hiện chuyển đổi ma trận xoay về ba góc quay Euler.

Cho một ma trận xoay 3x3:

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$$

Chuyển đổi qua 3 góc Euler:

$$\theta_x = \text{atan2}(r_{32}, r_{33})$$

$$\theta_y = \text{atan2}(-r_{31}, \sqrt{r_{32}^2 + r_{33}^2})$$

$$\theta_z = \text{atan2}(r_{21}, r_{11})$$

với θ_x, θ_z nằm trong khoảng từ $(-\pi, \pi)$ và θ_y chạy trong khoảng $(-\pi/2, \pi/2)$.

CHƯƠNG 3

THỰC NGHIỆM VÀ KẾT QUẢ

3.1 GIỚI THIỆU

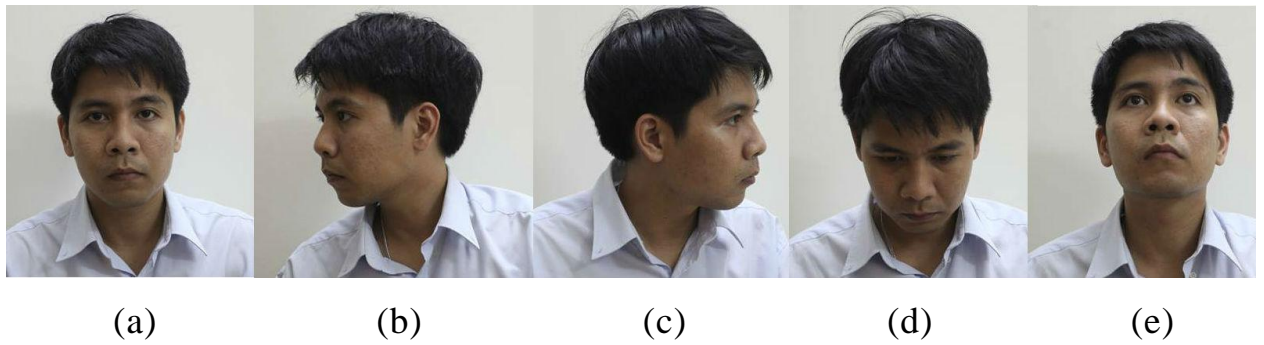
Như đã trình bày ở trên, bài toán nhận dạng hướng nhìn là một khía cạnh nhỏ trong bài toán nhận dạng khuôn mặt. Trên thực tế, các ảnh cần xử lý không phải là các ảnh mẫu chuẩn mà nó là những ảnh với nhiều tư thế khác nhau, trạng thái biểu cảm khác nhau của khuôn mặt. Với kết quả nghiên cứu, tác giả đã khắc phục được một số khó khăn, nhược điểm của việc nhận diện khuôn mặt, giúp tăng độ chính xác cho việc nhận diện khuôn mặt. Trong báo cáo, tác giả đã trình bày một số phương pháp xác định hướng nhìn của mặt người trong ảnh là phương pháp xác định dựa trên các đặc trưng Haar và phương pháp xác định dựa trên các điểm mốc trên khuôn mặt, từ đó xây dựng chương trình, tiến hành thực nghiệm đánh giá kết quả.

3.2 DỮ LIỆU ĐẦU VÀO

3.2.1 Tư thế khuôn mặt người

Kỹ thuật xác định hướng mặt người trong ảnh tập trung vào việc phát hiện các tư thế, trạng thái khác nhau của khuôn mặt so với tư thế chuẩn (nhìn thẳng). Đối tượng cần phân tích là khuôn mặt một người trong tư thế chuẩn và các tư thế thay đổi. Dựa vào các thành phần được phát hiện trên khuôn mặt như mắt, mũi, miệng làm cơ sở phát hiện, xác định hướng của khuôn mặt và góc quay so với tư thế ban đầu.

Một khuôn mặt có rất nhiều tư thế khác nhau, tùy thuộc vào ba góc quay chính là Roll, Pitch và Yaw. Việc xác định được ba góc quay này sẽ giúp đưa ra tư thế của đầu người.



Hình 3.1. Khuôn mặt chuẩn và các tư thế thay đổi

- a. Khuôn mặt tư thế chuẩn
- b. Khuôn mặt hướng sang phải
- c. Khuôn mặt hướng sang trái
- d. Khuôn mặt hướng xuống dưới
- e. Khuôn mặt hướng lên trên

3.2.2 Tập ảnh đánh giá

Mục tiêu của đề tài là xác định được góc quay của khuôn mặt trong ảnh. Tập ảnh đầu vào là tập ảnh chuẩn được xây dựng với các góc quay đã được xác định bằng thực nghiệm. Các ảnh này sẽ được xử lý bằng chương trình được tác giả xây dựng và được tiến hành đánh giá, so sánh kết quả với số liệu chuẩn.

Tập cơ sở dữ liệu “Head Pose Image Database” có 2790 ảnh khuôn mặt đơn của 15 người với các góc xoay của khuôn mặt từ -90^0 đến 90^0 . Mỗi người có 2 series với 93 ảnh (93 tư thế khác nhau). Mục đích của việc có 2 series ảnh là để có thể huấn luyện và kiểm tra thuật toán trên các khuôn mặt đã biết và chưa biết. Các ảnh trong cơ sở dữ liệu bao gồm cả người đeo kính và không đeo kính, với nhiều màu da khác nhau. Bối cảnh của ảnh là đơn giản và tập trung vào các chi tiết trên khuôn mặt. Tất cả các ảnh được định dạng jpeg.

Các tập tin trong thư viện được xây dựng theo cú pháp sau:

person[Id][Serie][Number][Tilt][Pan].jpg

$Id = \{01, \dots, 15\}$

Số thứ tự của người,

$Serie = \{1, 2\}$

Số thứ tự series,

$Number = \{00, 01, \dots, 92\}$

Số thứ tự của file ảnh trong thư mục,

$Tilt = \{-90, -60, -30, -15, 0, +15, +30, +60, +90\}$

Góc quay theo phương ngang,

$Pan = \{-90, -75, -60, -45, -30, -15, 0, +15, +30, +45, +60, +75, +90\}$

Góc quay theo phương dọc.

Tương ứng mỗi ảnh có 1 tập tin .txt lưu thông tin chuẩn để đánh giá, tập tin đó được lưu với cú pháp:

person[Id][Serie][Number][Tilt][Pan].txt.

Bên trong có nội dung bao gồm:

[tập tin tương ứng hình ảnh]

Face

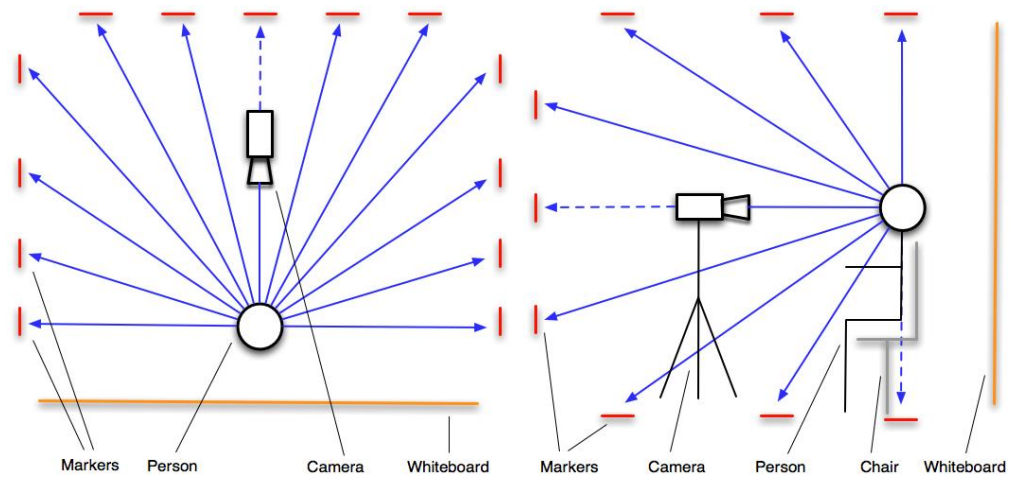
[Tọa độ trung tâm khuôn mặt X]

[Tọa độ trung tâm khuôn mặt Y]

[Độ rộng khuôn mặt]

[Độ cao khuôn mặt]

Các ảnh được chụp bằng cách đặt các dấu mốc trong phòng chụp, mỗi dấu mốc ứng với mỗi tư thế xoay theo hai chiều (pan, tilt). Khoảng cách từ người tới máy ảnh là 2m. Người mẫu sẽ nhìn lần lượt vào 93 dấu mốc trong căn phòng là sẽ được chụp 93 ảnh tương ứng.



Hình 3.2. Phương pháp tạo thư viện ảnh

Dưới đây là một vài hình ảnh trong thư viện chuẩn được tác giả sử dụng, thư viện này gồm tập hợp nhiều ảnh với nhiều tư thế khác nhau của đầu người.



Hình 3.3. Tư thế của khuôn mặt khi quay ngang
thay đổi góc α



Hình 3.4. Tư thế của khuôn mặt khi cúi xuống, ngẩng lên



Hình 3.5. Các tư thế của khuôn mặt
khi ngẩng lên trên và quay từ phải qua trái



Hình 3.6. Các tư thế của khuôn mặt khi nhìn xuống từ trái qua phải

3.3CHƯƠNG TRÌNH THỬ NGHIỆM

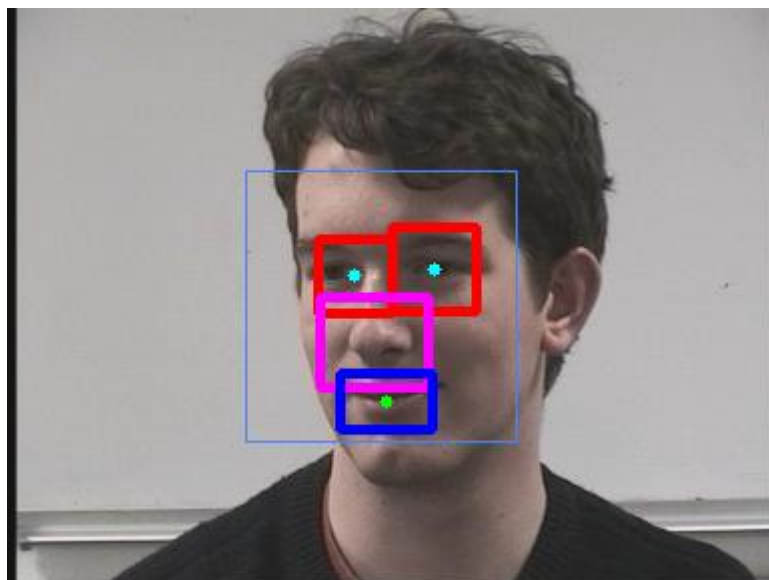
Trong chương 2, tác giả đã trình bày ba kỹ thuật phát hiện hướng mặt người trong ảnh: phát hiện hướng mặt người trong ảnh dựa trên tâm mắt, dựa theo các đặc trưng Haar và dựa vào các điểm mốc trên khuôn mặt cùng ma trận xoay.

Kỹ thuật phát hiện hướng mặt người dựa trên tâm mắt và dựa theo các đặc trưng Haar tuy đơn giản nhưng còn nhiều tồn tại và hạn chế. Kỹ thuật này chỉ làm việc với mô hình 2D của khuôn mặt, không xác định được chính xác vị trí, kích thước của các chi tiết trên khuôn mặt như mắt, mũi, miệng và đặc biệt là không xác định đúng đường bao trên

khuôn mặt dẫn đến các công thức tính toán góc xoay không còn chuẩn xác, cho kết quả sai đặc biệt với góc quay lớn.



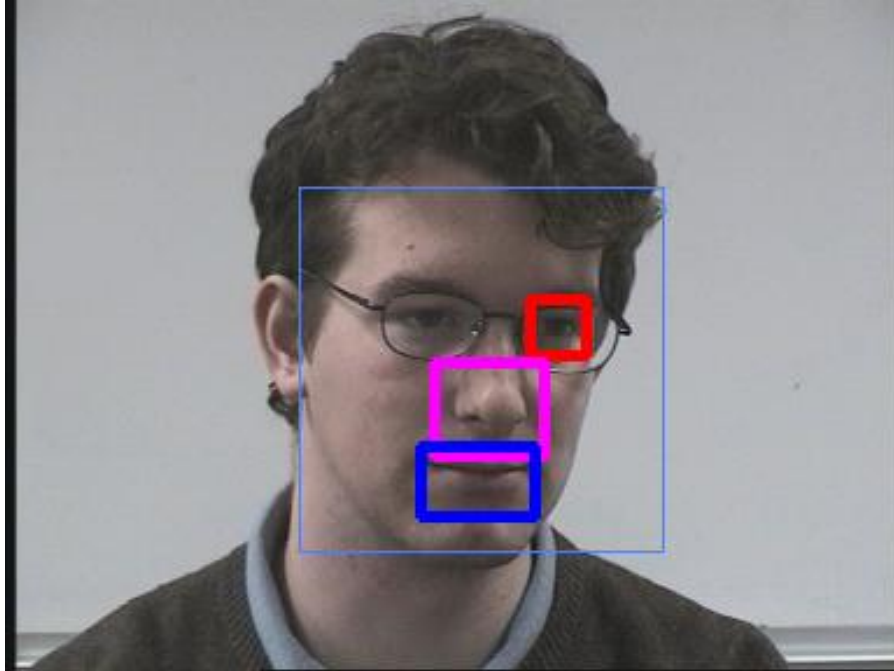
Hình 3.7. Không xác định được khuôn mặt



Hình 3.8. Xác định sai góc do xác định sai đường bao của khuôn mặt

Một số trường hợp, phương pháp dựa trên đặc trưng Haar nhận diện không đầy đủ các chi tiết trên khuôn mặt, dẫn đến việc không đưa ra được kết quả về góc nghiêng của khuôn mặt. Ngược lại phương pháp

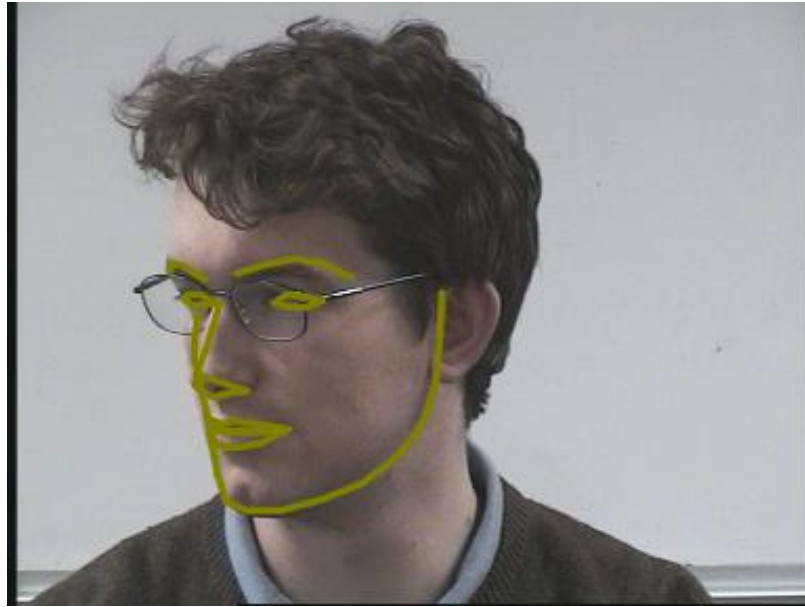
dựa trên điểm mốc chứng tỏ ưu thế ở khả năng nhận diện tốt ngay cả với góc quay lớn.



Hình 3.9. Phương pháp dùng đặc trưng Haar không nhận diện đủ chi tiết trên khuôn mặt



Hình 3.10. Phương pháp xác định điểm mốc nhận dạng chính xác



Hình 3.11. Phương pháp nhận dạng điểm mốc ở góc quay lớn

Ngược lại, kỹ thuật phát hiện hướng mặt người dựa theo các điểm mốc trên khuôn mặt cùng ma trận xoay có nhiều ưu điểm hơn.

Ngoài ra, việc sử dụng các thuật toán, thư viện hoàn chỉnh đã giúp nâng hiệu suất quá trình nhận dạng và phát hiện hướng khuôn mặt.

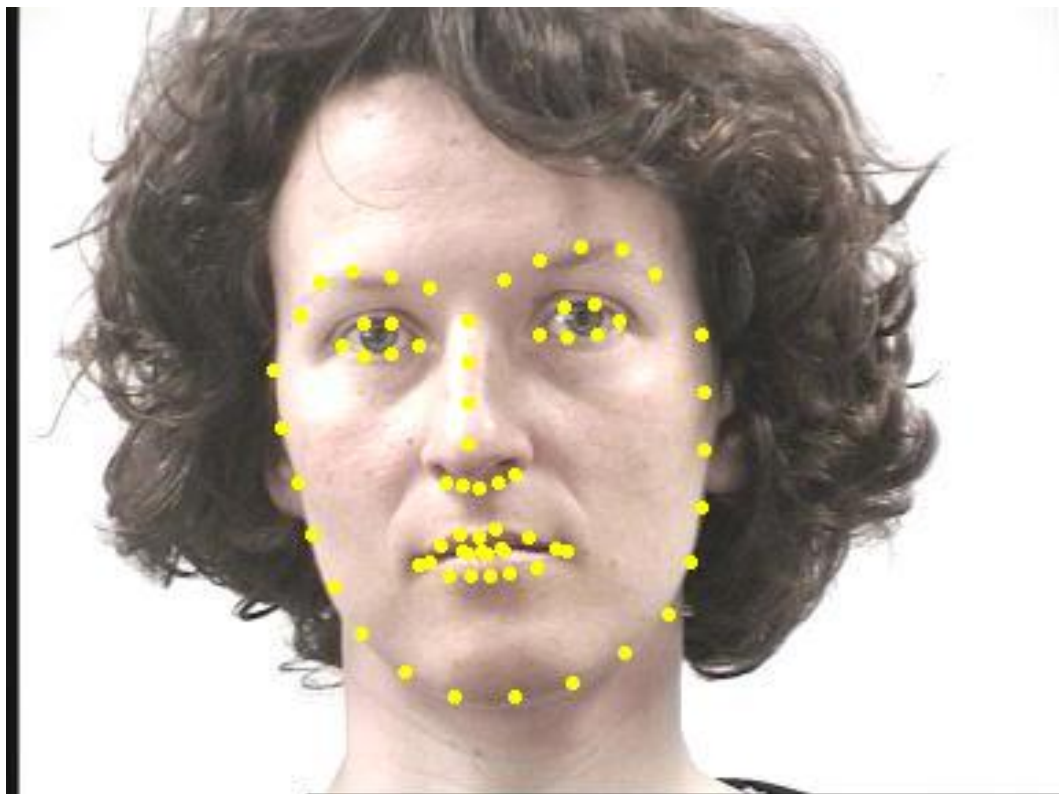
Kỹ thuật này cho phép xác định đầy đủ và chính xác các góc quay của khuôn mặt. Do vậy, tác giả đã xây dựng chương trình thử nghiệm theo kỹ thuật phát hiện hướng mặt người dựa theo các điểm mốc trên khuôn mặt cùng ma trận xoay. Trong quá trình xây dựng chương trình thử nghiệm, tác giả sử dụng thư viện mã nguồn mở OpenCV của Intel và thư viện Dlib.

Trong đó, OpenCV (Open Source Computer Vision) là một thư viện mã nguồn mở về thị giác máy với hơn 500 hàm và hơn 2500 các thuật toán đã được tối ưu về xử lý ảnh, và các vấn đề liên quan tới thị giác máy. OpenCV được thiết kế một cách tối ưu, sử dụng tối đa sức mạnh của các dòng chip đa lõi... để thực hiện các phép tính toán trong thời gian thực, nghĩa là tốc độ đáp ứng của nó có thể đủ nhanh cho các

ứng dụng thông thường. Và Dlib là một bộ công cụ chứa các thuật toán máy học và công cụ cho việc tạo ra các phần mềm phức tạp trong C++ để giải quyết các vấn đề thế giới thực. Đặc biệt là cung cấp công cụ hỗ trợ mạnh mẽ cho lĩnh vực xử lý ảnh như chuyển đổi không gian màu sắc tự động giữa các loại điểm ảnh khác nhau, phát hiện cạnh và hình thái đối tượng trong ảnh, các thuật toán trích xuất đặc trưng SURF, HOG, và FHOG, công cụ để phát hiện các đối tượng trong hình ảnh bao gồm nhận diện khuôn mặt phía trước và ước lượng tư thế đối tượng

Chương trình thử nghiệm thực hiện tìm góc quay của khuôn mặt trong hình ảnh đầu vào.

Đầu tiên ta sẽ thực hiện phát hiện khuôn mặt, tìm các điểm mốc để xác định đường bao khuôn mặt, các bộ phận mắt, mũi miệng.

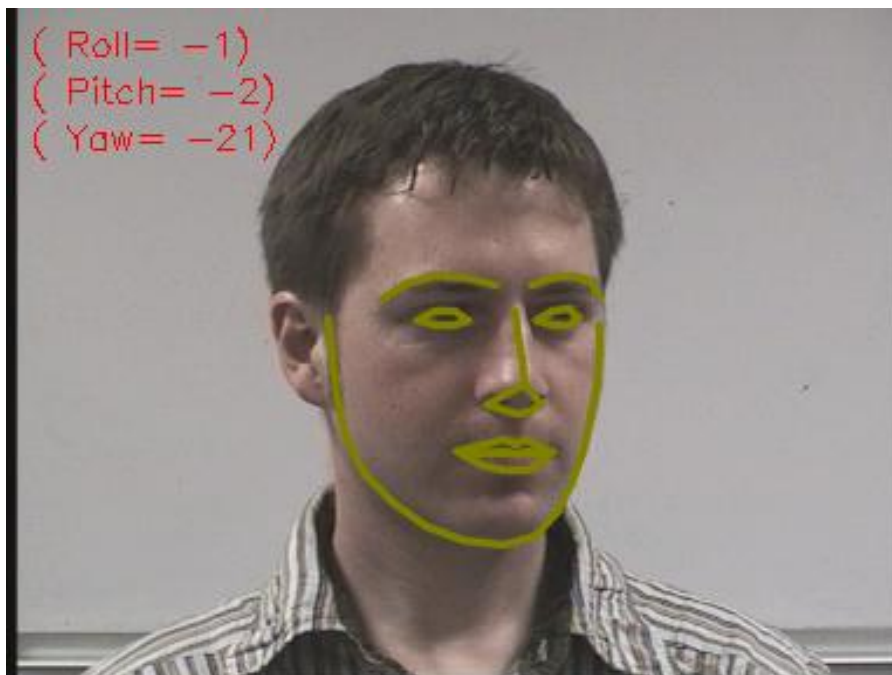


Hình3.12. Các điểm mốc trên khuôn mặt



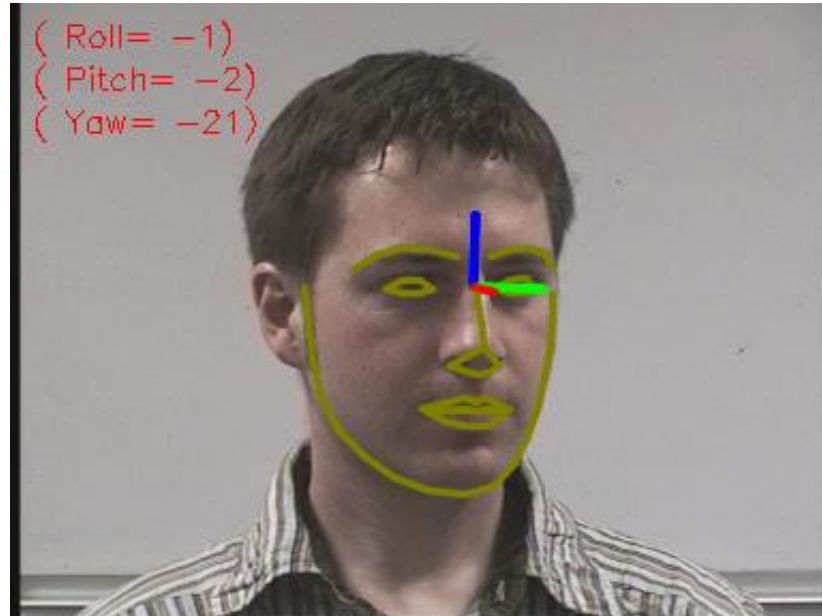
Hình3.13. Đường bao khuôn mặt và các bộ phận trên mặt

Từ các điểm mốc ta sẽ tìm được ma trận xoay, chuyển hóa ma trận xoay ta sẽ có được góc quay khuôn mặt theo 3 trục x(roll), y(pitch), z(yaw).



Hình3.14. Góc xoay của khuôn mặt

Bên cạnh đó, ta có thể biểu diễn được góc nhìn khuôn mặt theo hướng vector



Hình 3.15. Góc xoay của khuôn mặt biểu diễn hướng Vector

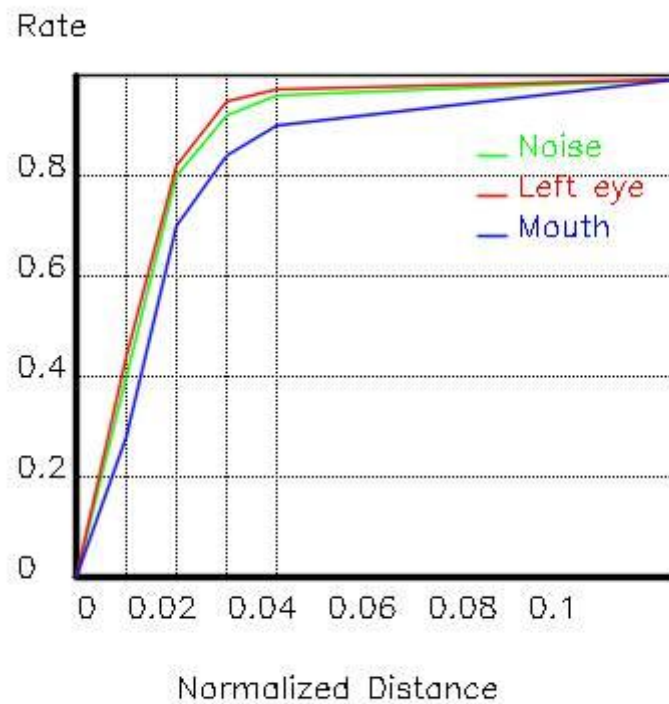
Ngoài ra, để đánh giá phương pháp ước lượng góc quay của khuôn mặt dựa trên các điểm mốc, ta thực hiện đánh giá dựa trên 2 tiêu chí.:

Thứ nhất là độ chính xác của bộ phát hiện các thành phần của khuôn mặt. Tác giả đánh giá tính chính xác của bộ phát hiện khuôn mặt của các cá nhân trên bộ cơ sở dữ liệu HeadPoseImage Database gồm các hình ảnh về khuôn mặt phía trước và nghiêng trái, phải,... Mỗi hình ảnh trong tập kiểm tra có một khuôn mặt duy nhất. Để đánh giá hiệu quả hoạt động của bộ phát hiện các thành phần của khuôn mặt, chúng ta tính tỷ lệ phát hiện đó được cho bởi công thức:

$$\text{Tỷ lệ phát hiện} = \frac{\text{số khuôn mặt được xác định đúng}}{\text{số hình ảnh dùng để kiểm tra}}$$

Để tính toán tỷ lệ phát hiện, chúng ta cần định nghĩa một tiêu chuẩn trong việc phát hiện đúng các thành phần khuôn mặt. Ở đây tác giả sử dụng khoảng cách Euclide giữa vị trí của chi tiết được phát hiện

và vị trí thật sự. Để làm cho khoảng cách này độc lập với kích thước hình ảnh, khoảng cách này được chuẩn hóa bởi các khung giới hạn.



Hình 3.16. Hiệu suất phát hiện

Có thể thấy, với ngưỡng khoảng cách là 0.02 thì tỷ lệ phát hiện đạt đến 80%, lên đến hơn 90% tại ngưỡng 0.04.

Thứ hai là về ước lượng tư thế khuôn mặt, tác giả thực hiện đánh giá trên cùng bộ dữ liệu trên, sử dụng mô hình 3D của khuôn mặt được cung cấp bởi thư viện dlib. Để đánh giá hiệu suất, tác giả thực hiện tính toán sai số trung bình tuyệt đối cho 3 góc quay.

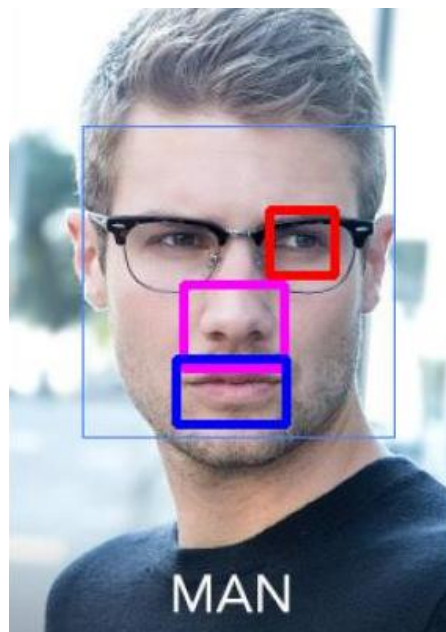
	Roll	Pitch	Yaw
Frontal	1.8 °	2.65 °	3.77 °
Others	10.22 °	13.5 °	8.1 °

Hình3.17. Hiệu suất ước lượng

Có thể thấy, các khuôn mặt với tư thế nhìn thẳng được xấp xỉ với độ chính xác cao, trong khi các tư thế nghiêng quá nhiều sẽ đem lại sai số cao hơn.

3.4 MỘT SỐ TRƯỜNG HỢP ĐẶC BIỆT

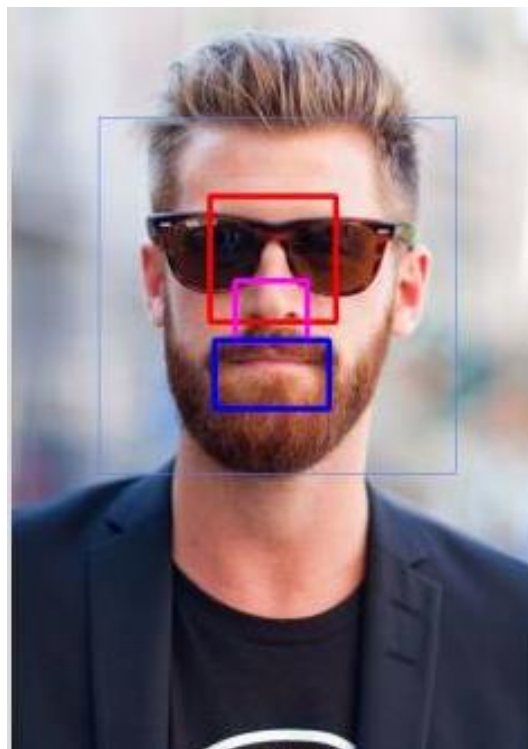
Trong quá trình nhận dạng nhằm xác định hướng nhìn của khuôn mặt, một số thành phần lạ có thể xuất hiện như: Mũ, kính mắt, ria mép... hoặc một số thành phần bị che khuất bởi các thành phần khác trên khuôn mặt. Với sự xuất hiện hoặc biến mất của các thành phần này, quá trình nhận dạng khuôn mặt cũng sẽ gặp nhiều khó khăn. Khuôn mặt người có thể được nhận dạng không đầy đủ các thành phần như mắt, mũi, miệng hoặc rút trích các đặc trưng này không chính xác từ đó dẫn đến việc đưa ra góc quay không chính xác. Tác giả thực nghiệm và đánh giá hai phương pháp, với phương pháp dựa trên các đặc trưng Haar, việc xác định các chi tiết gặp nhiều khó khăn. Ngược lại phương pháp điểm mốc lại cho kết quả khá chính xác.



Hình 3.18. Sử dụng phương pháp đặc trưng Haar đối với ảnh người đeo kính



Hình 3.19. Sử dụng phương pháp điểm mốc với ảnh người có đeo kính



Hình 3.20. Phương pháp đặc trưng Haar

với ảnh đeo kính và có râu



Hình 3.21. Phương pháp điểm mốc với ảnh đeo kính và có râu



Hình 3.22. Phương pháp điểm mốc với ảnh chụp nhiều người

Việc xác định chính xác các đặc điểm của khuôn mặt như mắt, mũi, miệng sẽ giúp kết quả của việc xác định góc quay của khuôn mặt chính xác hơn.

Phương pháp xác định khuôn mặt dựa trên điểm mốc cho thấy sự vượt trội về kết quả xác định, giúp nâng cao hiệu suất phát hiện so với phương pháp sử dụng đặc trưng Haar.

KẾT LUẬN

Nhận dạng khuôn mặt là một hướng nghiên cứu đã được quan tâm từ lâu và cũng đạt được nhiều thành tựu với nhiều ứng dụng cụ thể. Tuy nhiên việc nhận dạng khuôn mặt là một bài toán còn gặp nhiều khó khăn, thách thức, đặc biệt là việc xác định hướng xoay của khuôn mặt. Đa số các bài toán nhận dạng khuôn mặt chỉ cho kết quả đúng khi khuôn mặt ở tư thế thẳng đứng hoặc có góc quay nhỏ. Bài toán xác định hướng quay của khuôn mặt sẽ góp phần không nhỏ vào việc nhận dạng mặt người, giúp tăng khả năng phát hiện mặt người cũng như hỗ trợ nhiều ứng dụng cụ thể.

Trong đề tài, tác giả nghiên cứu các phương pháp xác định góc quay của khuôn mặt, cụ thể đã đạt được một số kết quả sau:

- Trình bày khái quát về xử lý ảnh, các phương pháp nhận dạng khuôn mặt, phương pháp nhận dạng các thuộc tính mắt – mũi – miệng trên khuôn mặt
- Tìm hiểu về một số phương pháp xác định góc quay của khuôn mặt như phương pháp “tâm hai mắt”, phương pháp xử dụng các đặc trưng Haar và đề xuất phương pháp dựa trên các điểm mốc để nâng cao hiệu quả của việc xác định hướng quay của khuôn mặt.
- Xây dựng chương trình thử nghiệm, đánh giá các phương pháp trên, đưa ra kết quả là góc quay theo các chiều của khuôn mặt.

Trên cơ sở tìm hiểu về các kỹ thuật nhận dạng khuôn mặt, tác giả đã xây dựng được chương trình xác định góc nghiêng của khuôn mặt theo hai phương pháp là phương pháp dựa trên các đặc trưng Haar và phương pháp xác định điểm mốc. Bằng kiểm chứng kết quả, tác giả nhận

thấy phương pháp xác định điểm mốc cho kết quả chính xác hơn, hiệu suất phát hiện cao hơn.

Đề tài mới chỉ dừng lại ở việc xây dựng chương trình đơn giản để đánh giá các phương pháp, chưa có các tính năng mở rộng cũng như chưa xây dựng được giao diện thân thiện người dùng.

Hướng phát triển của đề tài: Việc sử dụng thư viện Dlib kết hợp với phương pháp xác định điểm mốc trên khuôn mặt đã đem lại kết quả tốt hơn nhiều so với phương pháp dựa trên đặc trưng Haar. Hướng phát triển của đề tài là tiếp tục cải thiện hiệu suất phát hiện, đặc biệt đối với những trường hợp góc quay cúi xuống hoặc ngẩng lên. Áp dụng xây dựng các ứng dụng như phát hiện cảm xúc của người, phát hiện các hành vi bất thường thông qua xử lý ảnh.

TÀI LIỆU THAM KHẢO

- [1] E. Saber and A.M.Tekalp, “Frontal – View Face Detection and Facial Feature Extraction Using Color, Shape and Symmetry Based Cost Functions”, Pattern Recognition Letters, vol.17, no.8, pp.669-680, 1998.
- [2] Face pose estimation and its application in video shot selectionZ Yang, H Ai, B Wu, S Lao, L Cai - Pattern Recognition, 2004. ..., 2004 - ieeexplore.ieee.org
- [3] Head Pose Estimation in Computer Vision: A Survey Erik Murphy-Chutorian IEEE Transactions on Pattern Analysis and Machine Intelligence (Volume:31 , Issue: 4), 2008
- [4] Head pose estimation on depth data based on Particle Swarm Optimization by Pashalis Paderis, Xenophon Zabulis and Antonis A. Argyros
Institute of Computer Science - FORTHHeraklion, Crete, Greece
- [5] Li Wei, Eung Joo Lee, Multi-pose Face Recognition Using Head Pose Estimation and PCA Approach”, International Journal of Digital Content Technology and its Applications, 112~122,2010, Availabel at http://www.aicit.org/jdcta/ppl/jdcta4-1_revised2_paper12.pdf
- [6] M.Castrillón-Santana, O.Déniz-Suarez, L.Anton-Canalis, J.Lorenze-Navarro, “Face and Facial Feature Detection Evaluation, Institute of Intelligent Systems and Numerical Applications in Engineerin”, 2006, Availabel at <http://gias720.dis.ulpgc.es/Gias/Publications/visapp-2008-1.pdf>

- [7] M.Castrillón-Santana, O.Déniz-Suarez, L.Anton-Canalis, J.Lorenze-Navarro, “Face and Facial Feature Detection Evaluation, Institute of Intelligent Systems and Numerical Applications in Engineerin”, 2006,
- [8] M.Jones and P. Viola. Fast multi-view face detection. In CVPR 2003
- [9] Ming-Hsuan Yang, David J.Kriegman, and Narendra Ahuja, “Detecting Faces in Images: A Survey”, IEEE Transaction on Pattern Analysis and Machine Intelligent, vol. 24, no. 1, 2002.
- [10] N. Pears, T. Heseltine, and M. Romero. From 3D point clouds to pose-normalised depth maps. IJCV, 89(2-3):152–176, 2010.
- [11] One Millisecond Face Alignment with an Ensemble of Regression Trees by Vahid Kazemi and Josephine Sullivan, CVPR 2014
- [12] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, and N. Kumar. Localizing parts of faces using a consensus of exemplars. In CVPR 2011.
- [13] S. Malassiotis and M. Strintzis. Robust real-time 3D head pose estimation from range data. Pattern Recognition, 38:1153–1165, 2005
- [14] X. Zabulis, T. Sarmis, and A. A. Argyros. 3D head pose estimation from multiple distant views. In BMVC, 2009
- [15] Y. Tu, C. Zeng, C. Yeh, S. Huang, T. Cheng, and M. Ouhyoung. Real-time head pose estimation using depth map for avatar control. *CVGIP*, 2011

Trang Web

- [16] Availabel at <http://gias720.dis.ulpgc.es/Gias/Publications/visapp-2008-1.pdf>

- [17] Phillip Ian Wilson, John Fernandez, “Facial Feature Detection Using Haar Classifiers”, 2006, Available at :
<http://nichol.as/papers/Wilson/Facial%20feature%20detection%20using%20Haar.pdf>