

# MỤC LỤC

TÓM TẮT .....	4
CHƯƠNG 1: TỔNG QUAN VỀ PHÁT HIỆN KHUÔN MẶT .....	5
1.1. Giới thiệu .....	5
1.2. Tổng quan kiến trúc của một hệ thống nhận dạng mặt người .....	5
1.3. Một số lĩnh vực ứng dụng phát hiện khuôn mặt .....	6
1.4. Các hướng tiếp cận liên quan đến phát hiện và nhận dạng khuôn mặt .....	7
1.4.1. Hướng tiếp cận dựa trên tri thức .....	8
1.4.2. Hướng tiếp cận dựa trên các đặc trưng bất biến .....	11
1.4.3. Hướng tiếp cận dựa trên so khớp mẫu .....	13
1.4.4. Hướng tiếp cận dựa trên máy học (hay diện mạo) .....	16
1.5. Khó khăn và thách thức trong bài toán xác định khuôn mặt .....	18
CHƯƠNG 2: MỘT SỐ PHƯƠNG PHÁP MÁY HỌC .....	20
2.1. Khái niệm máy học .....	20
2.1.1. Các loại giải thuật trong máy học .....	20
2.1.2. Các chủ đề về máy học .....	21
2.2. Một số phương pháp máy học ứng dụng trong phát hiện khuôn mặt .....	22
2.2.1. Phương pháp Mạng neuron .....	22
2.2.2. Phương pháp SVM – support vector machine .....	23
2.2.3. Mô hình Makov ẩn .....	23
2.2.4. Mô hình Adaboost .....	24
CHƯƠNG 3: PHÁT HIỆN KHUÔN MẶT TRONG ẢNH DỰA VÀO MÀU DA .....	31
3.1. Tổng quan về các kỹ thuật nhận biết màu da dựa trên tính chất điểm ảnh .....	31
3.1.1. Giới thiệu .....	31
3.1.2. Không gian màu sử dụng cho mô hình hóa màu da .....	32
3.1.3. Mô hình hóa màu da .....	35
3.1.4. Mô hình hóa phân phổ màu da có tham số .....	38
3.1.5. So sánh kết quả các mô hình .....	39
3.1.6. Đánh giá phương pháp .....	41
3.1.7. Chọn lựa không gian màu và phương pháp mô hình hóa dùng để nhận biết màu da cho đề án .....	41
3.2. Nhận biết phân vùng màu da .....	42
3.2.1. Lọc khởi tạo .....	42
3.2.2. Nhận biết màu da .....	43
3.3. Trích chọn đặc trưng Haar - like .....	43
3.4. Huấn luyện dò tìm khuôn mặt .....	46
3.5. Quá trình dò tìm khuôn mặt .....	47

3.6. Hệ thống xác định vị trí khuôn mặt người.....	48
CHƯƠNG 4: CÀI ĐẶT ỨNG DỤNG .....	50
4.1. Môi trường TEST .....	50
4.2. Một số giao diện chính .....	50
4.3. Nhận xét.....	52
KẾT LUẬN.....	53
TÀI LIỆU THAM KHẢO .....	54

## **Lời cảm ơn**

Em xin chân thành gửi lời cảm ơn tới các thầy cô trường Đại học Dân lập Hải Phòng trong những năm vừa rồi đã dạy dỗ vun đắp kiến thức để em có điều kiện hoàn thành đồ án tốt nghiệp này.

Đặc biệt em xin cảm ơn thầy giáo Ngô Trường Giang khoa Công nghệ thông tin trường Đại học Dân lập Hải Phòng đã chỉ bảo tận tình giúp em hoàn thành đồ án tốt nghiệp.

Cuối cùng em xin gửi lời biết ơn đến gia đình, bạn bè đã ủng hộ và giúp đỡ em trong suốt thời gian qua. Do trình độ bản thân có hạn nên không tránh khỏi những thiếu sót, mong thầy cô và các bạn góp ý giúp đỡ để em có thể hoàn thiện đồ án của mình.

Em xin chân thành cảm ơn!

Hải Phòng, ngày tháng 10 năm 2010

Sinh viên thực hiện

Nguyễn Anh Đức

## TÓM TẮT

Phát hiện khuôn mặt là một vấn đề cơ bản trong ngành học quan sát bằng máy tính. Là một trong những giai đoạn của hệ thống nhận dạng mặt người cùng với nhiều ứng dụng rộng rãi và phổ biến khác như chỉ số hóa nội dung trong ảnh, hệ thống giám sát, hội thảo truyền hình..., phát hiện khuôn mặt đã và đang dành được sự quan tâm nghiên cứu của nhiều người trong suốt hai thập kỷ qua. Tuy nhiên, với những thách thức rất lớn gây ra do tính biến động của môi trường cũng như tính biến đổi cao của khuôn mặt, phát hiện mặt người vẫn đang là một trong những vấn đề mở đối với các nhà nghiên cứu. Và cho đến nay vẫn chưa có một phương pháp nào thực sự giải quyết hết các thách thức của phát hiện khuôn mặt. Nhiều cải tiến, nhiều đề xuất đã được đưa ra nhằm cải thiện chất lượng phát hiện khuôn mặt. Và một trong những cải tiến đó là dựa trên những đặc điểm bất biến hoặc có sự biến động có thể phân vùng được của khuôn mặt mà màu da của con người chính là một trong những đặc điểm đó.

Xuất phát từ ý tưởng kết hợp giữa màu da và một phương pháp phát hiện khuôn mặt hiệu quả, đề án đã tập trung xây dựng một hệ thống kết hợp giữa hệ thống phát hiện khuôn mặt dựa trên đặc trưng Haar và bộ lọc AdaBoost và phương pháp phân vùng màu da.

Phương pháp phân vùng màu da được kết hợp vào hệ thống như là một giới hạn cho phép giảm không gian tìm kiếm, tức là khuôn mặt sẽ chỉ được tìm kiếm trên các vùng màu da. Việc phân vùng này đã giúp cho thời gian phát hiện của hệ thống cải thiện đáng kể, hơn nữa nó còn giúp cho hệ thống hoạt động hiệu quả hơn.

Đề án với tên gọi là “**Một hướng tiếp cận trong phát hiện khuôn mặt trong ảnh**” được bố cục làm bốn mục chính.

Chương 1 : Giới thiệu, bao gồm phần giới thiệu, phần tổng quan kiến trúc của một hệ thống nhận dạng mặt người, phần một số ứng dụng liên quan tới phát hiện khuôn mặt và phần các hướng tiếp cận liên quan tới phát hiện và nhận dạng khuôn mặt. Chương 2 : Giới thiệu tổng quan về các phương pháp máy học. Chương 3 : Giới thiệu về các phương pháp tìm kiếm màu da trong ảnh. Chương 4 : Miêu tả kết quả việc kiểm thử chương trình. Đồng thời sẽ có những đánh giá và nhận xét về kết quả của chương trình.

# CHƯƠNG 1: TỔNG QUAN VỀ PHÁT HIỆN KHUÔN MẶT

## 1.1. Giới thiệu.

Hơn một thập kỷ qua có rất nhiều công trình nghiên cứu về bài toán xác định khuôn mặt người từ ảnh đen trắng, xám đến ảnh màu như ngày hôm nay. Các nghiên cứu đi từ bài toán đơn giản, mỗi ảnh chỉ có một khuôn mặt người nhìn thẳng vào thiết bị thu hình và đầu ở tư thế thẳng đứng trong ảnh đen trắng. Cho đến ngày hôm nay bài toán mở rộng cho ảnh màu, có nhiều khuôn mặt trong cùng một ảnh, có nhiều tư thế thay đổi trong ảnh. Không những vậy mà còn mở rộng cả phạm vi từ môi trường xung quanh khá đơn giản (trong phòng thí nghiệm) cho đến môi trường xung quanh rất phức tạp (như trong tự nhiên) nhằm đáp ứng nhu cầu của thực tế.

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, ...

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

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



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

Phát hiện khuôn mặt dò tìm và định vị những vị trí khuôn mặt xuất hiện trong ảnh hoặc trên các frame video. Phân đoạn khuôn mặt sẽ xác định vị trí mắt mũi, miệng, và các thành phần khác của khuôn mặt và chuyển kết quả này

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

### **1.3. Một số lĩnh vực ứng dụng phát hiện khuôn mặt.**

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

- Hệ thống phát hiện tội phạm: camera được đặt tại một số điểm công cộng như: siêu thị, nhà sách, trạm xe buýt, sân bay, v.v... Khi phát hiện được sự xuất hiện của các đối tượng là tội phạm, hệ thống sẽ gửi thông điệp về cho trung tâm xử lý.

- Hệ thống theo dõi nhân sự trong một đơn vị: giám sát giờ ra vào của từng nhân viên và chấm công.

- Hệ thống giao tiếp người máy: thay thế việc tương tác giữa người và máy theo những cách truyền thống như: bàn phím, chuột, v.v... Thay vào đó là sử dụng các giao tiếp trực quan: biểu cảm khuôn mặt, dấu hiệu, cử chỉ bằng tay (visual input, visual interaction).

- Hệ thống tìm kiếm thông tin trên ảnh, video dựa trên nội dung (chỉ mục theo người). Chẳng hạn như: đài truyền hình Việt Nam (VTV) có một kho dữ liệu video tin tức khá lớn cần tìm kiếm nhanh những đoạn video nào có G. Bush hoặc Bin Laden.

- Các hệ thống bảo mật dựa trên thông tin trắc sinh học: mặt người, vân tay, v.v... thay vì xác nhận mật khẩu, khóa, v.v...

## 1.4. Các hướng tiếp cận liên quan đến phát hiện và nhận dạng khuôn mặt

Có rất nhiều hướng tiếp cận trước đây đã thực hiện liên quan đến vấn đề phát hiện mặt người. Theo Ming-Hsuan Yang, có thể phân loại thành bốn hướng tiếp cận chính: dựa trên tri thức (knowledge-based), đặc trưng bất biến (feature invariant), đối sánh mẫu (template matching), và dựa vào diện mạo (appearance-based) phương pháp này thường dùng một mô hình máy học nên còn được gọi là phương pháp dựa trên máy học (machine learning-based).

**Các phương pháp dựa trên tri thức:** Hướng tiếp cận này chủ yếu dựa trên những luật được định nghĩa trước về khuôn mặt người. Những luật này thường là các mối quan hệ giữa các thành phần trên khuôn mặt. Có một số nghiên cứu từ rất sớm đã áp dụng phương pháp này như của Kanade 1973, G. Yang 1994, và Kotropoulos 1997.

**Hướng tiếp cận dựa trên các đặc trưng bất biến:** Hướng tiếp cận này cố gắng tìm kiếm những đặc trưng độc lập - những đặc trưng không phụ thuộc vào tư thế khuôn mặt, điều kiện chiếu sáng, và các khó khăn khác. Các đặc trưng như thế được gọi là bất biến và được sử dụng để phát hiện khuôn mặt. Những công trình sử dụng hướng tiếp cận này có thể kể như: K. C. Yow và R. Cipolla 1997, T. K. Leung 1995.

**Phương pháp 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. I. Craw 1992 đã áp dụng một mẫu cứng trong khi A. Lanitis 1995 sử dụng một mẫu có thể biến dạng trong bước phát hiện khuôn mặt.

**Phương pháp dựa trên máy học (hay 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 những 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 nhiều mô hình máy học được áp dụng trong hướng tiếp cận này: Eigenface (M. Turk và A. Pentland 1991 ), Mô hình dựa trên phân phối (K. K. Sung and T. Poggio 1998 ), Mạng Nơ-ron (H. Rowley 1998 ), Support Vector Machine (E. Osuna et al 1997 ), Phân lớp Bayes (H. Schneiderman và T. Kanade 1998), Mô hình

Markov ẩn (A. Rajagopalan et al 1998), và các mô hình tăng cường (AdaBoost của P. Viola và M. Jones 2001; FloatBoost do Stan Z. Li và Zhen Qiu Zhang 2004).

#### 1.4.1. Hướng tiếp cận dựa trên tri thức.

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

Một vấn đề khá phức tạp khi dùng hướng tiếp cận này là làm sao chuyển từ tri thức con người sang các luật một cách hiệu quả. Nếu các luật này quá chi tiết (chặt chẽ) thì khi xác định có thể xác định thiếu các khuôn mặt có trong ảnh, vì những khuôn mặt này không thể thỏa mãn tất cả các luật đưa ra. Nhưng các luật tổng quát quá thì có thể chúng ta sẽ xác định lầm một vùng nào đó không phải là khuôn mặt mà lại xác định là khuôn mặt. Và cũng khó khăn khi cần mở rộng yêu cầu của bài toán để xác định các khuôn mặt có nhiều tư thế khác nhau.

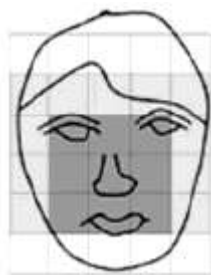


Hình 1-2: (a) Ảnh ban đầu có độ phân giải  $n=1$ ; (b), (c), và (d) Ảnh có độ phân giải  $n=4, 8, \text{ và } 16$ .

Yang và Huang dùng một phương thức theo hướng tiếp cận này để xác các khuôn mặt. Hệ thống của hai tác giả này bao gồm ba mức luật. Ở mức cao nhất, dùng một khung cửa sổ quét trên ảnh và thông qua một tập luật để tìm các ứng viên có thể là khuôn mặt. Ở mức kế tiếp, hai ông dùng một tập luật để mô tả tổng quát hình dáng khuôn mặt. Còn ở mức cuối cùng lại dùng một tập luật



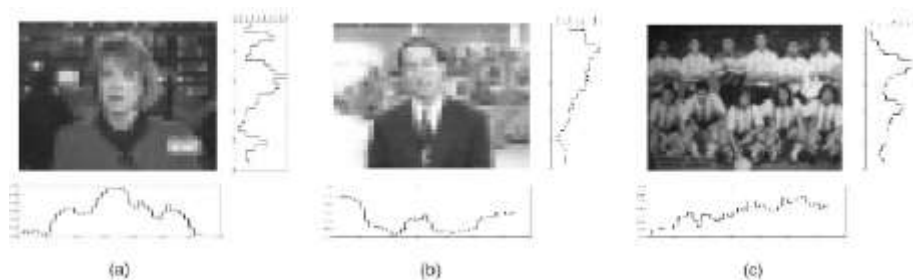
khác để xem xét ở mức chi tiết các đặc trưng khuôn mặt. Một hệ thống đa độ phân giải có thứ tự được dùng để xác định. Các luật ở mức cao nhất để tìm ứng viên như: “vùng trung tâm khuôn mặt có bốn phần với một mức độ đều cơ bản”, “phân xung quanh bên trên của một khuôn mặt có một mức độ đều cơ bản”, và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



Một loại tri thức của người nghiên cứu phân tích trên khuôn mặt.

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

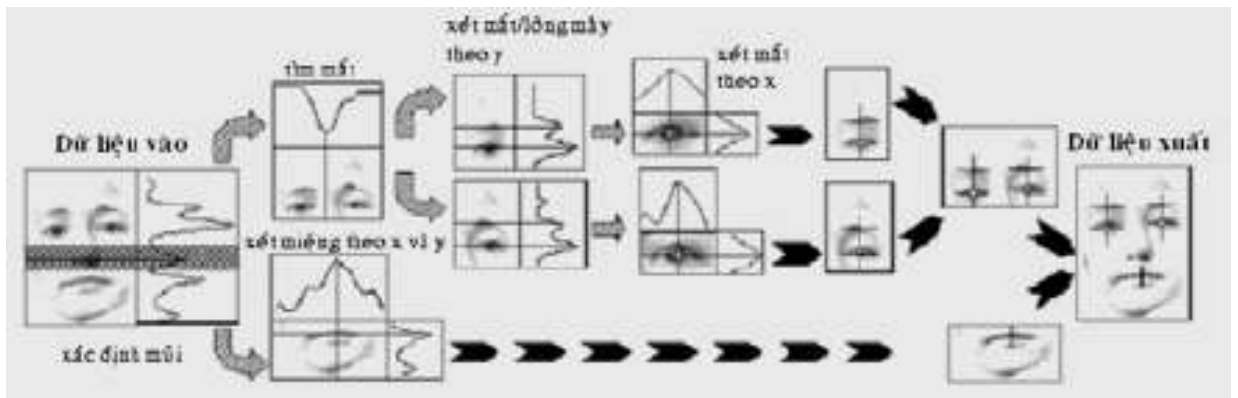
$$HI(x) = \sum_{y=1}^n I(x, y) \quad \text{và} \quad VI(y) = \sum_{x=1}^m I(x, y) \quad (1.1)$$



Hình 1-3: Phương pháp chiếu:

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

Dựa trên biểu đồ hình chiếu ngang, có hai cực tiểu địa phương khi hai ông xét quá trình thay đổi độ dốc của HI, đó chính là cạnh bên trái và phải của hai bên đầu. Tương tự với hình chiếu dọc VI, các cực tiểu địa phương cũng cho ta biết vị trí miệng, đỉnh mũi, và hai mắt. Các đặc trưng này đủ để xác định khuôn mặt. Hình 1-3.a cho một ví dụ về cách xác định như trên. Cách xác định này có tỷ lệ xác định chính xác là 86.5% cho trường hợp chỉ có một khuôn mặt thẳng trong ảnh và hình nền không phức tạp. Nếu hình nền phức tạp thì rất khó tìm, hình 1-3.b. Nếu ảnh có nhiều khuôn mặt thì sẽ không xác định được, hình 1-3.c.



Hình 1-4: Chiếu từng phần ứng viên để xác định khuôn mặt.

Mateos và Chicote dùng kết cấu để xác định ứng viên trong ảnh màu. Sau đó phân tích hình dáng, kích thước, thành phần khuôn mặt để xác định khuôn mặt. Khi tìm được ứng viên khuôn mặt, hai ông trích các ứng viên của từng thành phần khuôn mặt, sau đó chiếu từng phần này để xác thực đó có phải là thành phần khuôn mặt hay không, hình 1-4. Tỷ lệ chính xác hơn 87%.

Berber kết hợp mô hình màu da người và xác định cạnh để tìm ứng viên khuôn mặt người. Sau đó kết hợp quan hệ các đặc trưng và phương pháp chiếu các ứng viên khuôn mặt xuống hai trục: đứng và ngang để xác định ứng viên nào thật sự là khuôn mặt người.

## 1.4.2. Hướng tiếp cận dựa trên các đặc trưng bất biến

Đây là hướng tiếp cận theo kiểu bottom-up. Các tác giả cố gắng tìm các đặc trưng không thay đổi của khuôn mặt người để xác định khuôn mặt người. Dựa trên nhận xét thực tế, con người dễ dàng nhận biết các khuôn mặt và các đối tượng trong các tư thế khác nhau và điều kiện ánh sáng khác nhau, thì phải tồn tại các thuộc tính hay đặc trưng không thay đổi. Có nhiều nghiên cứu đầu tiên xác định các đặc trưng khuôn mặt rồi chỉ ra có khuôn mặt trong ảnh hay không. Các đặc trưng như: lông mày, mắt, mũi, miệng, và đường viền của tóc được trích bằng phương pháp xác định cạnh. Trên cơ sở các đặc trưng này, xây dựng một mô hình thống kê để mô tả quan hệ của các đặc trưng này và xác định sự tồn tại của khuôn mặt trong ảnh. Một vấn đề của các thuật toán theo hướng tiếp cận đặc trưng cần phải điều chỉnh cho phù hợp điều kiện ánh sáng, nhiễu, và bị che khuất. Đôi khi bóng của khuôn mặt sẽ tạo thêm cạnh mới, mà cạnh này lại rõ hơn cạnh thật sự của khuôn mặt, vì thế nếu dùng cạnh để xác định sẽ gặp khó khăn.

### 1.4.2.1. Các đặc trưng khuôn mặt

Sirohey đưa một phương pháp xác định khuôn mặt từ một ảnh có hình nền phức tạp. Phương pháp dựa trên cạnh (dùng phương pháp Candy và heuristics để loại bỏ các cạnh để còn lại duy nhất một đường bao xung quanh khuôn mặt. Một hình ellipse dùng để bao khuôn mặt, tách biệt vùng đầu và hình nền. Tỷ lệ chính xác của thuật toán là 80%.

Cũng dùng phương pháp cạnh như Sirohey, Chetverikov và Lerch dùng một phương pháp dựa trên blob và streak (hình dạng giọt nước và sọc xen kẽ), để xác định theo hướng các cạnh. Hai ông dùng hai blob tối và ba blob sáng để mô tả hai mắt, hai bên gò má, và mũi. Mô hình này dùng các streak để mô tả hình dáng ngoài của khuôn mặt, lông mày, và môi. Dùng ảnh có độ phân giải thấp theo biến đổi laplace để xác định khuôn mặt thông qua blob.

Graf đưa ra một phương pháp xác định đặc trưng rồi xác định khuôn mặt trong ảnh xám. Dùng bộ lọc để làm nổi các biên, các phép toán hình thái học (morphology) được dùng để làm nổi bật các vùng có cường độ cao và hình dáng chắc chắn (như mắt). Thông qua histogram để tìm các đỉnh nổi bật để xác định các ngưỡng chuyển ảnh xám thành hai ảnh nhị phân. Các thành phần dính nhau đều xuất hiện trong hai ảnh nhị phân thì được xem là vùng của ứng viên

khuôn mặt rồi phân loại xem có phải là khuôn mặt không. Phương pháp được kiểm tra trên các ảnh chỉ có đầu và vai của người. Tuy nhiên còn có một vấn đề ở đây là làm sao để sử dụng các phép toán morphology và làm sao xác định khuôn mặt trên các vùng ứng viên.

#### **1.4.2.2. Đặc trưng kết cấu**

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

Dai và Nakano dùng mô hình SGLD để xác định khuôn mặt người. Thông tin màu sắc được kết hợp với mô hình kết cấu khuôn mặt. Hai tác giả xây dựng thuật giải xác định khuôn mặt trong không gian màu, với các phần tựa màu cam để xác định các vùng có thể là khuôn mặt người. Một thuận lợi của phương pháp này là có thể xác định khuôn mặt không chỉ chụp thẳng và có thể có râu và có kính.

Mark và Andrew dùng phân bố màu da và thuật toán DoG (Difference of Gauss) để tìm các ứng viên, rồi xác thực bằng một hệ thống học kết cấu của khuôn mặt. Manian và Ross dùng biến đổi wavelet để xây dựng tập dữ liệu kết cấu của khuôn mặt trong ảnh xám thông qua nhiều độ phân giải khác nhau kết hợp xác suất thông kê để xác định khuôn mặt người. Mỗi mẫu sẽ có chín đặc trưng. Tỷ lệ chính xác là 87%, tỷ lệ xác định sai là 18%.

#### **1.4.2.3. Đặc trưng sắc màu của da**

Thông thường các ảnh màu không xác định trực tiếp trên toàn bộ dữ liệu ảnh mà các tác giả dùng tính chất sắc màu của da người (khuôn mặt người) để chọn ra được các ứng viên có thể là khuôn mặt người (lúc này dữ liệu đã thu hẹp đáng kể) để xác định khuôn mặt người.

#### 1.4.2.4. Đa đặc trưng

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

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

Sobottka và Pitas dùng các đặc trưng về hình dáng và màu sắc để xác định khuôn mặt người. Dùng một ngưỡng để phân đoạn trong không gian màu HSV để xác định các vùng có thể là màu da người (vùng giống màu da người). Các thành phần dính nhau sẽ được xác định bằng thuật toán tăng vùng ở độ phân giải thô. Xem xét tiền ứng viên nào vừa khớp hình dạng ellipse sẽ được chọn làm ứng viên của khuôn mặt. Sau đó dùng các đặc trưng bên trong như: mắt và miệng, được trích ra trên cơ sở các vùng mắt và miệng sẽ tốt hơn các vùng khác của khuôn mặt, sau cùng phân loại dựa trên mạng nơ-ron để biết vùng ứng viên nào là khuôn mặt người và vùng nào không phải khuôn mặt người. Tỷ lệ chính xác là 85%.

#### 1.4.3. Hướng tiếp cận dựa trên so khớp mẫu.

Trong so khớp mẫu, các mẫu chuẩn của khuôn mặt (thường là khuôn mặt được chụp thẳng) sẽ được xác định trước hoặc xác định các tham số thông qua một hàm. Từ một ảnh đưa vào, tính các giá trị tương quan so với các mẫu chuẩn về đường viền khuôn mặt, mắt, mũi và miệng. Thông qua các giá trị

tương quan này mà các tác giả quyết định có hay không có tồn tại khuôn mặt trong ảnh. Hướng tiếp cận này có lợi thế là rất dễ cài đặt, nhưng không hiệu quả khi tỷ lệ, tư thế, và hình dáng thay đổi (đã được chứng minh). Nhiều độ phân giải, đa tỷ lệ, các mẫu con, và các mẫu biến dạng được xem xét thành bất biến về tỷ lệ và hình dáng.

#### **1.4.3.1. Phương pháp đối sánh mẫu**

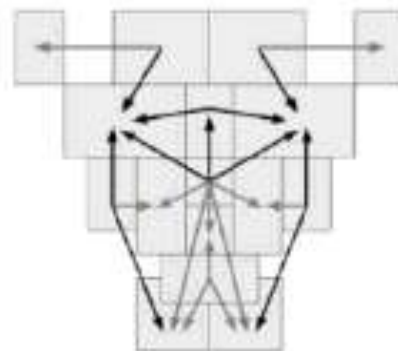
Sakai đã cố gắng thử xác định khuôn mặt người chụp thẳng trong ảnh. Ông dùng vài mẫu con về mắt, mũi, miệng, và đường viền khuôn mặt để mô hình hóa một khuôn mặt. Mỗi mẫu con được định nghĩa trong giới hạn của các đoạn thẳng. Các đường thẳng trong ảnh được trích bằng phương pháp xem xét thay đổi gradient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

Craw đưa ra một phương pháp xác định khuôn mặt người dựa vào các mẫu về hình dáng của các ảnh được chụp thẳng (dùng về bề ngoài của hình dáng khuôn mặt). Đầu tiên dùng phép lọc Sobel để tìm các cạnh. Các cạnh này sẽ được nhóm lại theo một số ràng buộc. Sau đó, tìm đường viền của đầu, quá trình tương tự được lặp đi lặp lại với mỗi tỷ lệ khác nhau để xác định các đặc trưng khác như: mắt, lông mày, và môi. Sau đó Craw mô tả một phương thức xác định dùng một tập có 40 mẫu để tìm các đặc trưng khuôn mặt và điều khiển chiến lược dò tìm.

Sinha dùng một tập nhỏ các bất biến ảnh trong không gian ảnh để mô tả không gian các mẫu ảnh. Tư tưởng chính của ông dựa vào sự thay đổi mức độ sáng của các vùng khác nhau của khuôn mặt (như hai mắt, hai má, và trán), quan hệ về mức độ sáng của các vùng còn lại thay đổi không đáng kể. Xác định các cặp tỷ số của mức độ sáng của một số vùng (một vùng tối hơn hay sáng hơn) cho ta một lượng bất biến khá hiệu quả. Các vùng có độ sáng đều được xem như một mẫu tỷ số mà là mẫu thô trong không gian ảnh của một khuôn mặt với độ thích hợp ít dùng để chọn như các đặc trưng chính của khuôn mặt như hai mắt, hai má, và trán. Lưu giữ thay đổi độ sáng của các vùng trên khuôn mặt trong một tập thích hợp với các cặp quan hệ sáng hơn – tối hơn giữa các

vùng nhỏ. Một khuôn mặt được xác định khi một ảnh thỏa tất cả các cặp sáng hơn – tối hơn. Ý tưởng này xuất phát từ sự khác biệt của cường độ giữa các vùng kề cận bộ, sau này được mở rộng trên cơ sở biến đổi wavelet để biểu diễn cho xác định người đi bộ, xác định xe hơi, xác định khuôn mặt. Ý tưởng của Sinha còn được áp dụng cho hệ thống thị giác của robot. Hình 1-5 cho thấy mẫu nổi bật trong 23 quan hệ được định nghĩa. Dùng các quan hệ này để phân loại, có 11 quan hệ thiết yếu (các mũi tên màu đen) và 12 quan hệ xác thực (các mũi tên xám). Mỗi mũi tên là một quan hệ. Một quan hệ thỏa mãn mẫu khuôn mặt khi tỷ lệ giữa hai vùng vượt qua một ngưỡng và 23 quan hệ này vượt ngưỡng thì xem như xác định được một khuôn mặt.

Phương pháp so khớp mẫu theo thứ tự để xác định khuôn mặt người do Miao trình bày. Ở giai đoạn đầu tiên, ảnh sẽ được xoay từ  $-20^\circ$  đến  $20^\circ$  với mỗi bước là  $5^\circ$  và theo thứ tự. Xây dựng ảnh đa độ phân giải, hình 1-1, rồi dùng phép toán Laplace để xác định các cạnh. Một mẫu khuôn mặt gồm các cạnh mô tả sáu thành phần: hai lông mày, hai mắt, một mũi, và một miệng. Sau đó áp dụng heuristic để xác định sự tồn tại của khuôn mặt trong ảnh, phương pháp này cho phép xác nhiều khuôn mặt, nhưng kết quả không tốt bằng xác định một khuôn mặt (chụp thẳng hoặc xoay) trong ảnh xám.



Hình 1-5: mẫu khuôn mặt, có 16 vùng và 23 quan hệ (các mũi tên).

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

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

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

#### **1.4.3.2. Các mẫu bị biến dạng**

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

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

#### **1.4.4. Hướng tiếp cận dựa trên máy học (hay diện mạo).**

Trái ngược với các phương pháp so khớp mẫu với các mẫu đã được định nghĩa trước bởi những chuyên gia, các mẫu trong hướng tiếp cận này được học từ các ảnh mẫu. Một cách tổng quát, các phương pháp theo hướng tiếp cận này áp dụng các kỹ thuật theo hướng xác suất thống kê và máy học để tìm những đặc tính liên quan của khuôn mặt và không phải là khuôn mặt. Các đặc tính đã được học ở trong hình thái các mô hình phân bố hay các hàm biệt số nên dùng có thể dùng các đặc tính này để xác định khuôn mặt người. Đồng thời, bài toán giảm số chiều thường được quan tâm để tăng hiệu quả tính toán cũng như hiệu quả xác định. Có nhiều phương pháp áp dụng xác suất thống kê để giả quyết. Một ảnh hay một vector đặc trưng xuất phát từ một ảnh được xem như một biến ngẫu nhiên  $x$ , và biến ngẫu nhiên có đặc tính là khuôn mặt hay không phải khuôn mặt bởi công thức tính theo các hàm mật độ phân lớp theo điều kiện  $p(x | \text{khuôn mặt})$  và  $p(x | \text{không phải khuôn mặt})$ . Có thể dùng phân loại Bayes



hoặc khả năng cực đại để phân loại một ứng viên là khuôn mặt hay không phải là khuôn mặt. Không thể cài đặt trực tiếp phân loại Bayes bởi vì số chiều của  $x$  khá cao, bởi vì  $p(x | \text{khuôn mặt})$  và  $p(x | \text{không phải khuôn mặt})$  là đa thức, và chưa thể hiểu nếu xây dựng các dạng tham số hóa một cách tự nhiên cho  $p(x | \text{khuôn mặt})$  và  $p(x | \text{không phải khuôn mặt})$ . Có khá nhiều nghiên cứu theo hướng tiếp cận này quan tâm xấp xỉ có tham số hay không có tham số cho  $p(x | \text{khuôn mặt})$  và  $p(x | \text{không phải khuôn mặt})$ . Các tiếp cận khác trong hướng tiếp cận dựa trên diện mạo là tìm một hàm biệt số (như: mặt phẳng quyết định, siêu phẳng để tách dữ liệu, hàm ngưỡng) để phân biệt hai lớp dữ liệu: khuôn mặt và không phải khuôn mặt. Bình thường, các mẫu ảnh được chiếu vào không gian có số chiều thấp hơn, rồi sau đó dùng một hàm biệt số (dựa trên các độ đo khoảng cách) để phân loại, hoặc xây dựng mặt quyết định phi tuyến bằng mạng nơ-ron đa tầng. Hoặc dùng SVM (Support Vector Machine) và các phương thức kernel, chiếu hoàn toàn các mẫu vào không gian có số chiều cao hơn để dữ liệu bị rời rạc hoàn toàn và ta có thể dùng một mặt phẳng quyết định phân loại các mẫu khuôn mặt và không phải khuôn mặt.

#### **1.4.4.1. Eigenface.**

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

Turk và Pentland áp dụng PCA để xác định và nhận dạng khuôn mặt. Tương tự, dùng PCA trên tập huấn luyện ảnh các khuôn mặt để sinh các ảnh riêng (còn gọi là eigenface) để tìm một không gian con (không gian khuôn mặt) trong không gian ảnh. Các ảnh khuôn mặt được chiếu vào không gian con này và được gom nhóm lại. Tương tự các ảnh không có khuôn mặt dùng để huấn luyện cũng được chiếu vào cùng không gian con và gom nhóm lại. Các ảnh khi chiếu vào không gian khuôn mặt thì không bị thay đổi tính chất cơ bản, trong khi chiếu các ảnh không có khuôn mặt thì xuất hiện sự khác nhau cũng không ít. Xác định sự có mặt của một khuôn mặt trong ảnh thông qua tất cả khoảng cách giữa các vị trí trong ảnh và không gian ảnh. Khoảng cách này dùng để xem xét có hay không có khuôn mặt người, kết quả khi tính toán các khoảng cách sẽ cho ta một bản đồ về khuôn mặt. Có thể xác định được từ cực tiểu địa phương của bản đồ này. Có nhiều nghiên cứu về xác định khuôn mặt, nhận dạng, và trích đặc trưng từ ý tưởng vector riêng, phân rã, và gom nhóm. Sau đó Kim phát triển cho ảnh màu, bằng cách phân đoạn ảnh để tìm ứng để không gian tìm kiếm bớt đi.

### **1.5. Khó khăn và thách thức trong bài toán xác định khuôn mặt.**

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

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

- Nhiều khuôn mặt có vùng da dính lẫn nhau.

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

- Tỷ lệ xác định chính xác là tỷ lệ số lượng các khuôn mặt người được xác định đúng từ hệ thống khi sử dụng một phương pháp để xây dựng so với số lượng khuôn mặt người thật sự có trong các ảnh (detection rate).
- Số lượng xác định nhầm là số lượng vùng trong ảnh không phải là khuôn mặt người mà hệ thống xác định nhầm là khuôn mặt người (false positives).
- Thời gian thực hiện là thời gian để máy tính xác định khuôn mặt người trong ảnh (running time).

## CHƯƠNG 2: MỘT SỐ PHƯƠNG PHÁP MÁY HỌC

### 2.1. Khái niệm máy học

Học máy, có tài liệu gọi là Máy học, (tiếng Anh: *machine learning*) là một lĩnh vực của trí tuệ nhân tạo liên quan đến việc phát triển các kĩ thuật cho phép các máy tính có thể "học". Cụ thể hơn, máy học là một phương pháp để tạo ra các chương trình máy tính bằng việc phân tích các tập dữ liệu. Máy học có liên quan lớn đến thống kê, vì cả hai lĩnh vực đều nghiên cứu việc phân tích dữ liệu, nhưng khác với thống kê, máy học tập trung vào sự phức tạp của các giải thuật trong việc thực thi tính toán. Nhiều bài toán suy luận được xếp vào loại bài toán NP-khó, vì thế một phần của máy học là nghiên cứu sự phát triển các giải thuật suy luận xấp xỉ mà có thể xử lí được.

Máy học có tính ứng dụng rất cao bao gồm máy truy tìm dữ liệu, chẩn đoán y khoa, phát hiện thẻ tín dụng giả, phân tích thị trường chứng khoán, phân loại các chuỗi DNA, nhận dạng tiếng nói và chữ viết, dịch tự động, chơi trò chơi và cử động rô-bốt (*robot locomotion*).

Một số hệ thống máy học nỗ lực loại bỏ nhu cầu trực giác của con người trong việc phân tích dữ liệu, trong khi các hệ thống khác hướng đến việc tăng sự cộng tác giữa người và máy. Không thể loại bỏ hoàn toàn tác động của con người vì các nhà thiết kế hệ thống phải chỉ định cách biểu diễn của dữ liệu và những cơ chế nào sẽ được dùng để tìm kiếm các đặc tính của dữ liệu. Máy học có thể được xem là một nỗ lực để tự động hóa một số phần của phương pháp khoa học. Một số nhà nghiên cứu máy học tạo ra các phương pháp bên trong các khuôn khổ của thống kê Bayes.

#### 2.1.1. Các loại giải thuật trong máy học

Các thuật toán máy học được phân loại theo kết quả mong muốn của thuật toán. Các loại thuật toán thường dùng bao gồm:

- Học có giám sát -- trong đó, thuật toán tạo ra một hàm ánh xạ dữ liệu vào tới kết quả mong muốn. Một phát biểu chuẩn về một việc học có giám sát là bài toán phân loại: chương trình cần học (cách xấp xỉ biểu hiện của) một hàm ánh xạ một vector tới một vài lớp bằng cách xem xét một số mẫu dữ\_liệu - kết\_quả của hàm đó.

- Học không giám sát -- mô hình hóa một tập dữ liệu, không có sẵn các ví dụ đã được gắn nhãn.
- Học nửa giám sát -- kết hợp các ví dụ có gắn nhãn và không gắn nhãn để sinh một hàm hoặc một bộ phân loại thích hợp.
- Học tăng cường -- trong đó, thuật toán học một chính sách hành động tùy theo các quan sát về thế giới. Mỗi hành động đều có tác động tới môi trường, và môi trường cung cấp thông tin phản hồi để hướng dẫn cho thuật toán của quá trình học.
- Chuyển đổi -- tương tự học có giám sát nhưng không xây dựng hàm một cách rõ ràng. Thay vì thế, cố gắng đoán kết quả mới dựa vào các dữ liệu huấn luyện, kết quả huấn luyện, và dữ liệu thử nghiệm có sẵn trong quá trình huấn luyện.
- Học cách học -- trong đó thuật toán học thiên kiến quy nạp của chính mình, dựa theo các kinh nghiệm đã gặp.

Phân tích hiệu quả các thuật toán máy học là một nhánh của ngành thống kê, được biết với tên lý thuyết học điện toán.

### 2.1.2. Các chủ đề về máy học

Mô hình hóa các hàm mật độ xác suất điều kiện: hồi quy và phân loại

- Mạng nơ-ron
- Cây quyết định
- Lập trình biểu thức gen
- Lập trình di truyền
- Hồi quy quá trình Gauss
- Phân tích biệt thức tuyến tính
- k láng giềng gần nhất
- Độ dài thông điệp tối thiểu
- Cảm tri nguyên
- Hàm cơ sở xuyên tâm
- Máy hỗ trợ vector

Mô hình hóa các hàm mật độ xác suất qua các mô hình phát sinh:

- Thuật toán cực đại kì vọng
- Các mô hình đồ họa gồm mạng Bayes và mạng Markov
- Ánh xạ topo phát sinh

Các kỹ thuật suy luận xấp xỉ đúng:

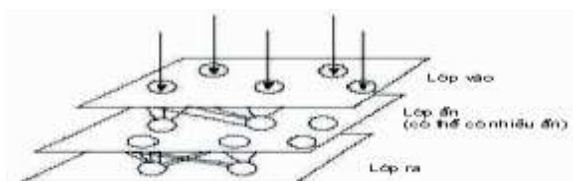
- Chuỗi Markov phương pháp Monte Carlo
- Phương pháp biến thiên

Tối ưu hóa: hầu hết các phương pháp trên đều sử dụng tối ưu hóa hoặc là các thể hiện của các thuật toán tối ưu hóa.

## 2.2. Một số phương pháp máy học ứng dụng trong phát hiện khuôn mặt

### 2.2.1. Phương pháp Mạng neuron

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



Hình 2-1: Mô hình mạng Neuron

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

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

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

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

### **2.2.2. Phương pháp SVM – support vector machine**

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

### **2.2.3. Mô hình Markov ẩn**

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

## 2.2.4. Mô hình Adaboost

### 2.2.4.1. Hướng tiếp cận theo AdaBoost

Phương pháp dò tìm AdaBoost dựa trên ý tưởng xây dựng các bộ dò tìm yếu mặc dù độ chính xác không cao nhưng có thời gian xử lý rất nhanh. Tuy nhiên khi kết hợp các bộ dò tìm lại có thể đạt độ chính xác cao.

Phương pháp AdaBoost sử dụng kết hợp các đặc trưng vốn dĩ tính toán rất nhanh, thích hợp cho việc dò tìm trong thời gian thực.

Các bộ phân loại AdaBoost có thể xây dựng phân tầng với độ phức tạp xử lý từ thấp đến cao, nhằm loại nhanh các ứng viên xấu (không phải mặt người) vốn dĩ nhiều hơn nhiều các ứng viên là mặt người để cho bộ phân loại phức tạp hơn (chỉ còn lại ít ứng viên chưa bị loại).

### 2.2.4.2. Thuật toán ADABOOST

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

$$H_M(x) = \sum_{m=1}^M h_m(x) \quad (2.1)$$

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

$$J(H_M) = \sum e^{-y_i H_M(x_i)} \quad (2.2)$$

Thuật toán AdaBoost xây dựng hàm  $h(x)$  bằng cách giảm tối đa (2.2).



Cho

$$H_{M-1}(x) = \sum_{m=1}^{M-1} h_m(x) \quad (2.3)$$

$H_M(x)$  tốt nhất cho phân loại mạnh  $H_M(x) = H_{M-1}(x) + h_m(x)$

Là hàm dẫn tới giá trị nhỏ nhất:

$$H_m = \arg \min J(H(x) h(x)) \quad (2.4)$$

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

$$h_M(x) = \frac{1}{2} \log \frac{P(y=+1|x, \omega^{(M-1)})}{P(y=-1|x, \omega^{(M-1)})} \quad (2.5)$$

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

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

$$L_M(x) = \frac{1}{2} \log \frac{P(x|y=+1, \omega)}{P(x|y=-1, \omega)} \quad (2.6)$$

$$T = \frac{1}{2} \left[ \log \frac{P(y=+1)}{P(y=-1)} \right] \quad (2.7)$$

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

Ta có một phương pháp để tính phương trình (2.7), ứng dụng khi học các bộ phân lớp tối ưu. Vì rút ra một bộ phân loại yếu trong một miền không gian nhiều chiều là công việc quan trọng, xin được đưa ra sau đây một mô hình thống kê học theo từng giai đoạn dựa trên vài đặc điểm vô hướng. Một đặc điểm vô hướng j của x được tính bằng một phép biến đổi từ không gian dữ liệu n chiều thành đường thẳng thực  $z_j(x) \in Z$ . Một đặc điểm có thể là hệ số, hay nói trong xử lý ảnh là phép biến đổi vi ba tín hiệu. Nếu phương pháp tìm kiếm ước lượng được sử dụng như phép biến đổi  $z_j(x)$  đơn giản được xem là tọa độ thứ j của x. Một danh sách K đặc điểm ứng cử viên có thể được tạo  $Z = \{z_1(x), \dots, z_K(x)\}$ . Trong phần sau, chúng ta sử dụng  $z(m)$  để biểu diễn cho đặc điểm được chọn trong giai đoạn m, và  $z_k(x)$  là đặc điểm được tính toán từ x sử dụng phép biến đổi thứ k.

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

$$p(x|y, \omega^{M-1}) \approx p(z_{(1)}, z_{(2)}, \dots, z_{(M-1)}, z_k, |y, \omega^{M-1}) \quad (2.8)$$

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

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

$$p(z_k, |y, z_{(1)}, z_{(2)}, \dots, z_{(M-1)}, \omega^{M-1}) \quad (2.9)$$

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

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

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

$$p(z_{(M-1)}|y, \omega^{(M-2)}) p(z_k|y, \omega^{(M-1)}) \quad (2.10)$$

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

Cho :

$$L_k^{(M)}(x) = \frac{P(z_k | y = +1, \omega^{(M-1)})}{P(z_k | y = -1, \omega^{(M-1)})} \quad \text{và} \quad h_k^{(M)}(x) = \frac{1}{2} L_k^{(M)}(x) - T \quad (2.11)$$

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

$$\Gamma^{(M)}(x) = h_k^{(M)}(x) | \forall k \quad (2.12)$$

### **Thuật toán AdaBoost**

Bước 0. Đầu vào

1. Tập  $Z = \{ (x_1, y_1), (x_2, y_2), \dots, (x_n, y_n) \}$  với:  
 $N = a + b$ ;  
 $a$  là số mẫu thuộc phân lớp  $y_i = +1$

2. b là số mẫu thuộc phân lớp  $y_i = -1$   
Số lớp yếu tối đa  $M_{\max}$  được kết hợp

Bước 1. Khởi tạo giá trị

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

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

$M = 0$

Bước 2. Suy diễn tiến

While  $M < M_{\max}$

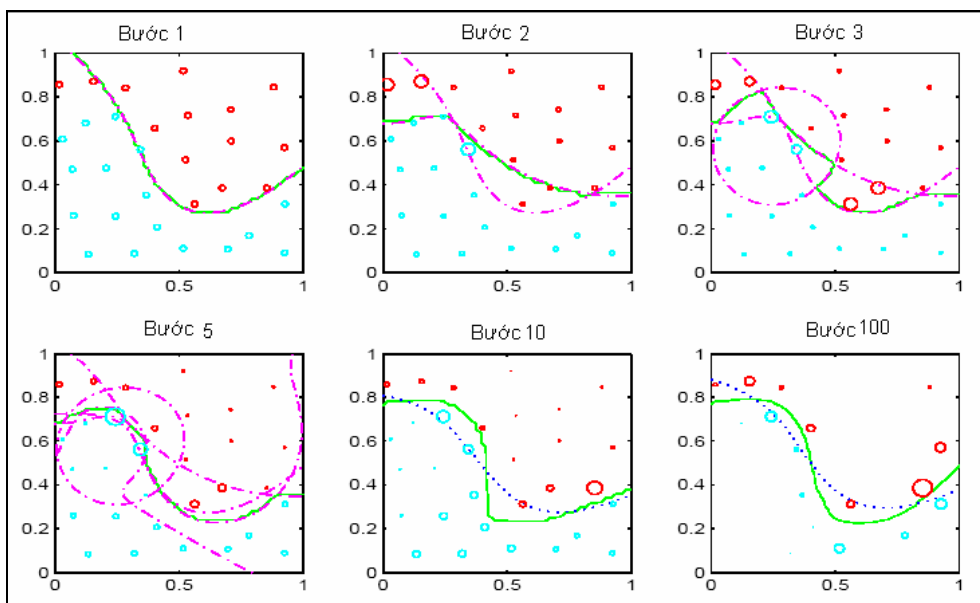
(1)  $M \leftarrow M + 1$

(2) Chọn  $h_m$  theo biểu thức (4)

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

Bước 3. Đầu ra

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

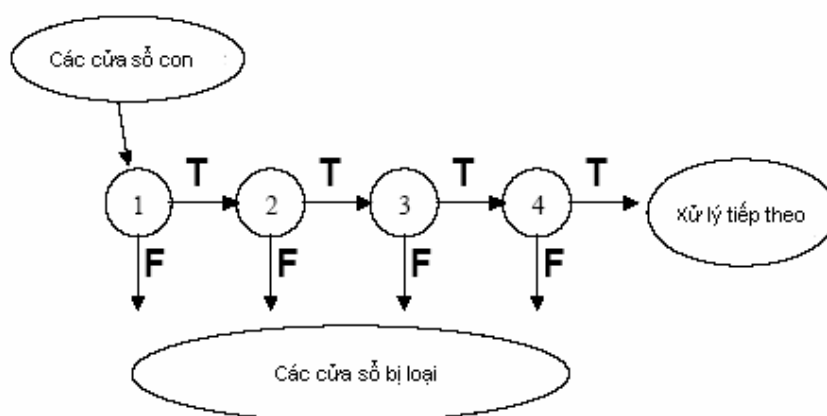


Hình 2-2: Ví dụ minh họa cho thuật toán AdaBoost

### 2.2.4.3. Bộ dò tìm phân tầng Adaboost

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

Nếu sử dụng bộ dò tìm kết hợp  $F = \{c_i\}$  với nhiều bộ dò tìm cơ bản  $f_i$  khác nhau cũng rơi vào tình trạng tương tự. Để có được độ chính xác cao, hoặc cần phải có số lượng lớn các bộ dò tìm, hoặc mỗi bộ dò tìm cần phải có nhiều đặc trưng, hoặc cả hai. Do đó cũng kéo theo tốc độ sẽ giảm. Một hướng khắc phục nhược điểm này là sử dụng bộ dò tìm phân tầng  $T = \{t_i\}$ . Bộ dò tìm phân tầng bao gồm nhiều tầng, mỗi tầng  $t_i = \{c_j\}$  là một bộ dò tìm kết hợp với số lượng các bộ dò tìm khác nhau nên có tốc độ và độ chính xác khác nhau. Khi dò tìm tất cả các khuôn mặt trong ảnh, tất cả các cửa sổ con  $W_0 = \{w_{i,j,s}\}$  với các kích thước  $s$  khác nhau tại các tọa độ  $(i,j)$  sẽ được kiểm tra xem có phải là mặt người hay không. Qua mỗi tầng  $t_i$ ,  $W_i = t_i(W_{i-1})$  trong đó  $|W_i| \ll |W_{i-1}|$  do qua mỗi tầng, các cửa sổ không phải ứng viên sẽ bị loại sớm. Điều này cho phép chúng ta xây dựng các tầng sao cho, càng về sau độ phức tạp (số lượng các bộ dò tìm cơ bản với các đặc trưng) càng lớn trong khi các tầng càng thấp thì độ phức tạp càng đơn giản và phải loại được nhiều ứng viên càng tốt nhưng tỷ lệ loại sai phải thấp.



Hình 2-3: Minh họa bộ dò tìm phân tầng

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

$$F = \prod_{i=1}^K f_i \quad (2.13)$$

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

$$D = \prod_{i=1}^K d_i \quad (2.14)$$

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

$$G = \prod_{i=1}^{|T|} D_i = \prod_{i=1}^{|T|} \prod_{j=1}^{|t_i|} c_j^i \quad (2.15)$$

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

Dưới đây là thuật toán huấn luyện một tầng với tỷ lệ loại sai  $f$  và độ chính xác  $d$  cho trước:

Bước 0. Đầu vào:

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

$d$  (tỷ lệ nhận đúng tối thiểu trong lớp)

$F_{target}$  (tỷ lệ nhận sai mẫu dương)

$P$ =tập mẫu dương

$N$ =tập mẫu âm

Bước 1. Khởi tạo:

$F_0=1.0$

$D_0=1.0$

$i=0$

Bước 2. Trong khi ( $F_i > F_{target}$ )

$i \leftarrow i + 1$

$n_i = 0; F_i = F_{i-1}$

Trong khi ( $F_i > f \times F_{i-1}$ )

$n_i = n_i + 1$

- Sử dụng  $P$  và  $N$  để huấn luyện một phân loại  $H$  với  $n_i$  đặc trưng, dùng Adaboost
- Thêm bộ phân loại hiện thời vào  $C$
- Tính  $F_i$  và  $D_i$  cho bộ phân loại  $C$  hiện thời trên tập hợp  $l$

- Giảm ngưỡng cho lớp thứ  $i$  cho đến khi bộ phân loại  $C$  hiện thời đạt tỉ lệ dò tìm tối thiểu là  $d \times D_{i-1}$  (điều này cũng ảnh hưởng  $F_i$ )
- $N \leftarrow \emptyset$
- Nếu  $F_i > F_{\text{target}}$  thì định giá bộ dò tìm  $C$  hiện thời trên tập ảnh không phải mặt người và đưa các mẫu dò tìm bị lỗi vào tập  $N$

Bước 3. Đầu ra: bộ dò tìm đa tầng  $C$

## **CHƯƠNG 3: PHÁT HIỆN KHUÔN MẶT TRONG ẢNH DỰA VÀO MÀU DA**

### **3.1. Tổng quan về các kỹ thuật nhận biết màu da dựa trên tính chất điểm ảnh**

#### **3.1.1. Giới thiệu**

Như đã trình bày trong phần trước, dựa vào màu sắc của da người cũng là một trong những phương pháp được dùng để phát hiện mặt người. Tuy nhiên nếu chỉ đơn thuần sử dụng màu sắc không thôi thì rất khó có thể đạt được hiệu quả cao trong phát hiện mặt người. Vì trong các khung cảnh thì có rất nhiều vật có màu sắc tương tự như màu của khuôn mặt. Tuy nhiên nếu kết hợp phương pháp này với các phương pháp khác lại có thể mang lại hiệu quả cao. Vì kinh nghiệm cho thấy màu da người có đặc tính màu riêng biệt, và đặc tính này cho phép dễ dàng nhận ra đâu là da người [8]. Và thông thường trong hướng tiếp cận phát hiện mặt người dựa trên thông tin xuất hiện trong ảnh, thì màu da được sử dụng như một bước phân vùng các vùng ảnh có màu sắc giống màu da, điều đó cho phép giảm không gian tìm kiếm khuôn mặt, cải thiện hiệu năng của hệ thống tìm kiếm. Do đó nhiều mô hình đã được xây dựng để có thể phát hiện được da người,

Khi xây dựng hay mô hình hóa một hệ thống phát hiện hay phân tách vùng màu da với mục đích sử dụng cho việc phát hiện mặt người, người ta thường đặt ra ba vấn đề chính. Thứ nhất là mô hình đó được xây dựng trong không gian màu nào, thứ hai là hàm phân phối của màu da được mô hình hóa chính xác đến mức độ nào và cuối cùng là sẽ xử lý vùng màu da được phân vùng cho nhận biết mặt người như thế nào. Trong phần này, sẽ chỉ đề cập đến hai câu hỏi trên, còn việc xử lý vùng da như thế nào cho việc phát hiện mặt người, sẽ đề cập đến trong chương sau, với một phương pháp cụ thể được chọn để sử dụng cho bài đồ án này.

Phương pháp được đề cập trong phần này là phương pháp phát hiện da người dựa trên đặc tính điểm ảnh, nghĩa là sẽ phân lớp điểm ảnh thành hai lớp, lớp là lớp điểm ảnh có thuộc màu da và lớp kia không phải là màu da. Các điểm ảnh là hoàn toàn độc lập với nhau. Ngược lại với phương pháp này là phương pháp dựa trên đặc tính vùng ảnh.

Phát hiện màu da dựa trên đặc tính điểm ảnh có một lịch sử phát triển khá dài, tuy nhiên trong khuôn khổ phần tổng quan này, chỉ đề cập và so sánh những kỹ thuật đã được công bố và được đánh giá hiệu quả.

Mục đích cuối cùng của phần tổng quan này là thu tập các kỹ thuật đã được công bố, mô tả những ý tưởng chính của kỹ thuật đó, tổng hợp và đưa ra những ưu điểm, nhược điểm và những đặc trưng của từng kỹ thuật. Từ đó sẽ đưa ra quyết định lựa chọn kỹ thuật phù hợp dùng để phân vùng màu da áp dụng cho bài đồ án này.

### **3.1.2. Không gian màu sử dụng cho mô hình hóa màu da**

Trong lĩnh vực đo màu, cũng như các lĩnh vực trong truyền tín hiệu hình ảnh và video sử dụng rất nhiều không gian màu với các tính chất khác nhau. Và trong số đó nhiều không gian màu được áp dụng cho vấn đề mô hình hóa màu da. Sau đây là tóm lược nhóm các không gian màu được sử dụng rộng rãi nhất cũng như các tính chất của chúng

#### **3.1.2.1. Không gian màu RGB**

RGB là không gian màu cơ bản được áp dụng từ lâu cho màn hình CRT. Trong không gian màu này, mỗi điểm màu là sự kết hợp của ba thành phần đơn màu (Đỏ - Red, Xanh lá cây- Green và xanh da trời : Blue). Đây là một trong những không gian màu được sử dụng phổ biến nhất cho việc xử lý và lưu trữ dữ liệu ảnh số. Tuy nhiên do tính tương quan cao giữa các kênh, giá trị cảm nhận không đồng nhất, sự pha trộn giữa dữ liệu thành phần màu và dữ liệu về độ sáng mà không gian RGB không được ưa thích sử dụng cho việc phân tích màu cũng như trong các thuật toán nhận dạng dựa trên màu sắc.

#### **3.1.2.2. Không gian RGB chuẩn hóa**

Không gian RGB chuẩn hóa là không gian màu nhận được từ không gian RGB cơ bản theo công thức chuẩn hóa đơn giản sau đây:

$$r = \frac{R}{R+G+B}; g = \frac{G}{R+G+B}; b = \frac{B}{R+G+B} \quad (3.1)$$

Có thể dễ dàng thấy rằng, trong không gian này,  $r+g+b=1$ . Do đó chỉ cần hai trong ba thành phần trên là đủ để biểu diễn không gian màu này, thành phần thứ ba sẽ không còn giá trị và có thể được bỏ qua, để rút ngắn được số chiều của không gian này. Hai thành phần còn lại thường được gọi là các thành



phần “màu tinh khiết” (“pure colors”). Thông thường, hai thành phần r và b thường được giữ lại, còn g bị rút bỏ đi. Tính chất cần chú ý của không gian màu này đó là tính bất biến đối với của bề mặt. Nghĩa là, nếu như không quan tâm đến ánh sáng xung quanh, thì không gian chuẩn hóa RGB là bất biến đối với sự thay đổi về hướng bề mặt liên quan đến nguồn chiếu (tất nhiên là dưới một vài giả thiết nhất định). Kết hợp với phép chuyển đổi đơn giản từ không gian màu RGB cơ bản mà không gian RGB chuẩn hóa này ngày càng được sử dụng rộng rãi trong nhiều lĩnh vực, trong đó có lĩnh vực nhận dạng.

### 3.1.2.3. HIS, HSV, HSL - Độ bão hòa của màu

Không gian màu dựa trên tính bão hòa màu được giới thiệu khi có những nhu cầu trong việc xác định số lượng tính chất màu. Chúng miêu tả màu sắc với những giá trị thuộc về trực giác, dựa trên ý kiến của các họa sỹ về những trạng thái khác nhau của màu sắc, trạng thái bão hòa cũng như từng tông màu khác nhau. Hue biểu thị cho màu trời (như màu đỏ, màu xanh lá cây, màu đỏ tía và màu vàng) của một vùng ảnh, saturation (độ bão hòa) là thước đo cho giới mức ngưỡng màu của một vùng ảnh. Các khái niệm như “intensity” (cường độ), “lightness” (tính dịu) hay “value”(giá trị) liên quan đến độ sáng của màu. Giá trị trực giác của các thành phần trong không gian màu này và sự phân biệt rõ ràng giữa độ sáng với cá thành phần màu của không gian màu là ưu điểm mà giúp cho không gian này được sử dụng phổ biến trong vấn đề phân vùng màu da.

Công thức chuyển từ không gian RGB sang không gian này như sau:

$$H = \arccos \frac{1/2((R - G) + (R - B))}{\sqrt{((R - G)^2 + (R - B)(G - B))}}$$

$$S = 1 - 3 \frac{\min(R, G, B)}{R + G + B} \quad (3.2)$$

$$V = \frac{1}{3}(R + G + B)$$

Ngoài ra còn có thể tính Hue và Saturation bằng cách sử dụng hàm log cho các thành phần màu của không gian màu RGB. Phương pháp này có thể làm giảm sự độc lập của các thành phần màu theo mức sáng.

Hệ tọa độ cực giữa Hue và Saturation có thể gây ra nhiều khó khăn trong mô hình màu da, chính vì vậy người ta còn chuyển nó sang hệ tọa độ Đề các theo công thức sau:

$$X = S \cos H, Y = S \sin H \quad (3.3)$$

#### 3.1.2.4. TSL – Tint, Saturation, Lightness(sắc thái, độ bão hòa, độ dịu )

Không gian chuẩn hóa thành phần màu và độ sáng TSL được chuyển từ không gian chuẩn hóa RGB theo công thức sau đây:

$$S = [9/5(r'^2 + g'^2)]^{1/2}$$

$$T = \begin{cases} \arctan(r'/g')/2\pi + 1/4 & (g' > 0) \\ \arctan(r'/g')/2\pi + 3/4 & (g' < 0) \\ 0 & (g' = 0) \end{cases} \quad (3.4)$$

$$L = 0.299R + 0.587G + 0.114B$$

Trong đó,  $r' = r - 1/3$ ,  $g' = g - 1/3$ .

Theo kết quả nghiên cứu, so sánh chín không gian màu sử dụng cho mô hình màu da thì không gian chuẩn hóa TSL thực hiện mô hình hóa màu da nhanh hơn các không gian khác.

#### 3.1.2.5. YcrCb

$Y_CrCb$  là không gian màu được sử dụng nhiều trong vấn đề nén ảnh. Màu sắc được biểu diễn bởi luma (đó là giá trị độ sáng được tính toán từ không gian RGB), gồm ba thành phần, một thành phần là tổng các trọng số từ RGB, hai thành phần màu khác nhau Cr và Cb được tạo ra bằng cách từ từ hai thành phần Red và Blue trong không gian màu RGB. Công thức để chuyển đổi như sau:

$$Y = 0.299R + 0.587G + 0.114B$$

$$Cr = R - Y \quad (3.5)$$

$$Cb = B - Y$$

Việc chuyển đổi đơn giản, tính phân chia rõ ràng của độ sáng và các thành phần màu là những đặc tính giúp cho không gian này lôi cuốn các nhà nghiên cứu sử dụng cho việc mô hình hóa màu da.

### 3.1.2.6. Các hệ tọa độ không gian màu khác

Bên cạnh YcrCb, một vài không gian màu khác được tạo ra từ chuyển đổi tuyến tính không gian RGB được sử dụng trong vấn đề phát hiện màu da. Như là YES, YUV hay YIQ. Tuy nhiên chúng ít được sử dụng hơn.

### 3.1.3. Mô hình hóa màu da

Mục đích cuối cùng của phát hiện màu da là xây dựng một quy tắc có tính quyết định. Đây là quy tắc sẽ giúp phân biệt một điểm ảnh là da hay không phải là da người. Thông thường, quy tắc này sẽ thiết lập một giá trị đo cho phép tính toán mức độ tương đồng giữa một điểm ảnh màu với đặc trưng màu da. Giá trị đo này được thiết lập như thế nào, công thức ra sao tùy thuộc vào từng phương pháp mô hình hóa màu da.

#### 3.1.3.1. Xác định ngưỡng cụ thể một điểm ảnh là màu da

Trong một số không gian màu, phương pháp xây dựng và xếp lớp màu da bằng cách xác định rõ ràng (thông qua một số quy tắc) biên giới các giá trị của điểm ảnh là màu da hay không. Ví dụ như:

##### **Trong không gian RGB:**

*(R,G,B) được xếp thuộc lớp màu da nếu như:*

$$R > 95 \text{ và } G > 40 \text{ và } B > 20 \text{ và}$$

$$\text{Max}(R,G,B) - \text{Min}(R,G,B) > 15 \text{ và}$$

$$|R-B| > 15 \text{ và } R > G \text{ và } R > B$$

Tính đơn giản của phương pháp này cũng thu hút nhiều sự tập trung nghiên cứu. Ưu điểm dễ thấy của phương pháp này đó là tính đơn giản của quy tắc nhận biết màu da. Điều này cho phép phân lớp một cách nhanh chóng và dễ dàng. Tuy nhiên kết quả đạt được khi phân lớp là không cao trong trường hợp tổng quát. Vì vậy khó khăn chính của phương pháp này nếu muốn có được hệ số nhận dạng cao đó là phải tìm ra được một không gian màu thích hợp cũng như các quy tắc tốt để nhận biết màu da trong không gian màu này.

Hiện nay người ta đang đề xuất sử dụng thuật toán máy học để tìm ra một không gian màu thích hợp cũng như các quy tắc phân lớp màu da với mong muốn có được hệ số nhận dạng cao. Tuy nhiên đó cũng chỉ mới là đề xuất và chưa có một kết quả cụ thể của một nghiên cứu nào được công bố.

Tuy nhiên, giữa và kết quả đạt được, chúng ta vẫn có thể tìm ra được những quy tắc cho phép nhận biết chắc chắn một điểm ảnh không phải là màu da. Những quy tắc này có thể được sử dụng làm bước lọc khởi tạo cho các phương pháp phân lớp phức tạp hơn giữa vùng màu da và vùng không phải màu da. Nó giúp cho quá trình phân lớp được thực hiện nhanh chóng hơn và đỡ tốn công hơn.

### **3.1.3.2. Phương pháp mô hình hóa màu da sử dụng phân phối không tham số**

Ý tưởng chính của phương pháp mô hình hóa màu da không tham số đó là ước lượng phân phối màu da từ dữ liệu huấn luyện mà không xuất phát từ một mô hình rõ ràng nào của màu da. Kết quả của phương pháp này thường được biểu diễn dưới dạng một bản đồ phân bố màu da (SPM – Skin Probability Map). Mỗi một giá trị phân bố được gán cho mỗi điểm trong không gian màu.

### **3.1.3.3. Bảng tra cứu chuẩn hóa (LUT – Lookup Table)**

Một số thuật toán phát hiện mặt người và bám sát mặt người sử dụng một lược đồ mức xám dựa trên hướng tiếp cận phân vùng các điểm ảnh là màu da. Không gian màu được lượng tự hóa thành từng nhóm, mỗi một nhóm đáp ứng cho một khoảng các thành phần màu. Các nhóm lược đồ này được tham chiếu tới một bảng gọi là bảng tra cứu. Mỗi một nhóm lưu trữ một số lượng lần xuất hiện của một màu khi tiến hành huấn luyện ảnh da người. Sau quá trình huấn luyện, biểu đồ sẽ tính toán và chuẩn hóa, chuyển sang giá trị biểu đồ trong phân phối xác suất miền rời rạc:

$$P_{skin}(c) = skin[c]/Norm \quad (3.6)$$

Trong đó,  $skin[c]$  nhận giá trị của nhóm lược đồ, đáp ứng cho véc tơ màu  $c$ ,  $Norm$  là một hệ số chuẩn hóa (tổng tất cả các giá trị của các nhóm biểu đồ) hay là giá trị lớn nhất của một nhóm biểu đồ. Giá trị chuẩn hóa của của bảng tra cứu các nhóm biểu đồ là căn cứ để cho phép quyết định một màu có là màu da hay không.

### **3.1.3.4. Phân lớp Bayes (Bayes Classifier)**

Giá trị của  $P_{skin}(c)$  trong công thức trên là một điều kiện xác suất –  $P(c|skin)$  – xác suất một màu quan sát  $c$  là một pixel màu da. Và xác suất thích

hợp được dùng để phát hiện màu da đó là  $P(\text{skin}|c)$  – xác suất quan sát màu được màu da khi xuất hiện một giá trị màu  $c$  rời rạc. Để tính giá trị này, ta sử dụng công thức Bayes quen thuộc:

$$P(\text{skin}|c) = \frac{P(c | \text{skin})P(\text{skin})}{P(c | \text{skin})P(\text{skin}) + P(c | -\text{skin})P(-\text{skin})} \quad (3.7)$$

Trong đó  $P(c|\text{skin})$  và  $P(c|-\text{skin})$  được tính trực tiếp từ biểu đồ màu da và không màu da. Xác suất toàn phần  $P(\text{skin})$  và  $P(-\text{skin})$  thì được ước lượng từ một số lượng các mẫu là màu da và không màu da trong tập mẫu huấn luyện. Bất đẳng thức  $P(\text{skin}|c) > \Theta$ , trong đó  $\Theta$  là một giá trị ngưỡng, có thể được sử dụng để trở thành quy tắc trong phát hiện màu da.

Công thức trên đôi khi hơi phức tạp, và để có thể tránh điều này, nếu như thực sự không cần phải biết một cách chính xác suất  $P(\text{skin}|c)$  và  $P(-\text{skin}|c)$  mà chỉ cần biết tỉ số giữa chúng thì người ta thường đưa về công thức như sau:

$$\frac{P(\text{skin} | c)}{P(-\text{skin} | c)} = \frac{P(c | \text{skin})P(\text{skin})}{P(c | -\text{skin})P(-\text{skin})} \quad (3.8)$$

So sánh công thức này với một ngưỡng có thể tạo ra một quy tắc cho phép phát hiện tỉ số màu da/không phải màu da. Sau một vài phép biến đổi, chúng ta nhận được công thức

$$\begin{aligned} \frac{P(\text{skin} | c)}{P(-\text{skin} | c)} &> \theta \\ \theta &= K \times \frac{1 - P(\text{skin})}{P(\text{skin})} \end{aligned} \quad (3.9)$$

Công thức trên có thể thấy rằng, việc chọn lựa giá trị của xác suất toàn phần không ảnh hưởng đến chất lượng của bộ phát hiện, vì với bất kì một xác suất toàn phần  $P(\text{skin})$  đều có thể chọn được một giá trị  $K$  phù hợp sao cho giá trị của ngưỡng là  $\theta$ .

### 3.1.3.5. Tổng kết phương pháp không tham số

Hai ưu điểm dễ thấy của phương pháp mô hình hóa phân phối không tham số đó là: thứ nhất, chúng có thể huấn luyện và sử dụng được một cách nhanh chóng. Thứ hai, chúng độc lập với lý thuyết về hình dạng của phân phối màu da (điều này không đúng trong mô hình hóa màu da có tham số). Tuy nhiên nhược điểm của phương pháp này đó là chúng yêu cầu nhiều bộ nhớ để lưu trữ và không có khả năng nội suy hay tạo ra dữ liệu huấn luyện. Lấy ví dụ như, chúng ta lượng tử hóa điểm ảnh trong không gian RGB về 8bit cho mỗi màu, khi đó

chúng ta phải cần một mảng có tới  $2^{24}$  phần tử để lưu trữ tập tất cả các xác suất của mô hình. Để có thể giảm bớt kích thước này bằng cách loại bỏ những dữ liệu huấn luyện nhỏ lẻ, không gian màu thường sử dụng kích thước  $128*128*128$ ,  $64*64*64$ ,  $32*32*32$ . Theo như nghiên cứu thì kích thước  $32*32*32$  là kích thước không gian mang lại hiệu quả cao nhất.

### 3.1.4. Mô hình hóa phân phối màu da có tham số

Hầu hết các mô hình màu da không tham số dựa trên biểu đồ xám đều yêu cầu rất nhiều bộ nhớ và hiệu năng của chúng phụ thuộc hoàn toàn của tập ảnh huấn luyện cố định. Vì vậy cần có một mô hình màu da có thể tự thêm hoặc tự tạo ra dữ liệu huấn luyện để, điều đó dẫn đến sự ra đời của mô hình phân phối tham số.

#### 3.1.4.1. Mô hình dựa trên phân phối Gaussian đơn.

Phân phối màu da có thể được mô hình hóa bởi phân phối Gaussian thêm vào hàm mật độ xác suất. Định nghĩa như sau:

$$P(c|skin) = \frac{1}{2\pi |\Sigma_s|^{1/2}} e^{-1/2(c-\mu_s)^t \Sigma_s^{-1}(c-\mu_s)} \quad (3.10)$$

Ở đây,  $c$  là một vectơ màu,  $\mu_s$  và  $\Sigma_s$  là hai tham số phân phối (véc tơ trung bình và ma trận hiệp phương sai).

Phương pháp mô hình hóa dựa trên phân phối đơn Gaussian đã được triển khai và nghiên cứu.

#### 3.1.4.2. Mô hình kết hợp dựa trên phân phối Gaussian

Một mô hình công phu, phức tạp hơn, có khả năng biểu diễn được phân phối phức tạp đó là mô hình kết hợp dựa trên phân phối Gaussian. Đây là mô hình mở rộng từ mô hình đơn Gaussian trên, trong trường hợp này, hàm phân phối mật độ xác suất là:

$$P(c|skin) = \sum_{i=1}^k \pi_i p_i(c | skin) \quad (3.11)$$

Trong đó,  $k$  là số lượng các thành phần được kết hợp,  $\pi_i$  là tham số kết hợp, thỏa mãn ràng buộc  $\sum_{i=1}^k \pi_i = 1$ , và  $p_i(c|skin)$  thỏa mãn hàm phối mật độ xác suất Gaussian, với mỗi véc tơ trung bình và ma trận hiệp phương sai của

nó. Huấn luyện mô hình được thực hiện với một kỹ thuật được biết đến nhiều gọi là thuật toán kì vọng tối đa (EM - Expectation Maximization), trong đó giả sử rằng số lượng các thành phần  $k$  là đã biết trước. Chi tiết việc huấn luyện mô hình kết hợp Gaussian với thuật toán EM này có thể được tìm thấy trong nhiều nghiên cứu. Việc phân lớp trong mô hình kết hợp Gaussian được thực hiện nhờ việc so sánh xác suất  $p(c|skin)$  với một vài giá trị ngưỡng.

Việc chọn lựa số lượng thành phần  $k$  ở đây là quan trọng. Vì nó ảnh hưởng đến độ chính xác của việc huấn luyện cho mô hình. Theo như những nghiên cứu hiện nay,  $k = 8$  là sự lựa chọn mang hiệu năng cao nhất cho mô hình kết hợp phân phối Gaussian.

### **3.1.4.3. Đa phân phối Gaussian**

Mức độ gần đúng của các nhóm màu da với phân phối Gaussian 3D trong không gian YcbCr đã được miêu tả trong nhiều bài báo. Một số lượng khác nhau các thuật toán phân nhóm K-trung bình được sử dụng cho nhóm Gaussian thực hiện việc huấn luyện mô hình. Các điểm ảnh được phân lớp thành lớp màu da nếu như khoảng cách Mahalanobis từ vectơ màu  $c$  đến trung tâm của cụm gần nhất trong mô hình nhỏ hơn một ngưỡng cho trước.

### **3.1.4.4. Tổng kết các phương pháp mô hình hóa theo tham số**

Tất cả các phương pháp mô hình hóa theo tham số được miêu tả như trên (ngoại trừ phương pháp 3.1.4.3) đều tính toán trên mặt phẳng các thành phần màu của không gian màu mà bỏ qua thông tin về độ sáng.

Dĩ nhiên, khi một mô hình phân phối cụ thể được sử dụng, sẽ có câu hỏi đặt ra về sự xác thực về giá trị của mô hình đó. Hiển nhiên, mô hình độc lập với hình dạng của phân phối trong không gian màu thì càng tốt hơn, do đó mô hình không tham số xét về mặt này hiển nhiên sẽ tốt hơn mô hình có tham số. Tuy nhiên do yêu cầu quá cao về bộ nhớ mà khi đánh giá hiệu năng thì mô hình có tham số lại có hiệu năng cao hơn. Điều này có thể thấy trong các bảng đánh giá ở nhiều bài báo.

### **3.1.5. So sánh kết quả các mô hình**

Để có thể đánh giá và so sánh hiệu năng của các phương pháp mô hình hóa màu da là không dễ, vì mỗi phương pháp thường được đề xuất của một nhóm các nhà nghiên cứu và được thử nghiệm trên cơ sở dữ liệu riêng. Và chưa

có một cơ sở dữ liệu chuẩn nào được công bố rộng rãi cho vấn đề này. Cơ sở dữ liệu huấn luyện và kiểm định được biết đến nhiều nhất đó là cơ sở dữ liệu của Compaq. Trong bảng so sánh được đưa ra dưới đây, là kết quả tốt nhất mà mỗi phương pháp đạt được, được tổng hợp bởi [Valimir Vezhnevets, Vassili Sazonov Alla Andreeva ], với kết quả thử nghiệm dựa trên cơ sở dữ liệu của Compaq nêu trên. Bảng sau đây sẽ so sánh hệ số phát hiện đúng và không đúng của từng phương pháp. Mặc dù các phương pháp là khác nhau về dữ liệu huấn luyện và tập dữ liệu test, cũng như chiến lược huấn luyện, bảng dưới đây vẫn mô tả một bức tranh toàn cảnh về hiệu năng của các phương pháp

<b>Phương Pháp</b>	<b>Nhận biết đúng</b>	<b>Nhận biết sai</b>
Bayes SPM trong RGB (Jones và Regh 1999)	<b>80%</b> <b>90%</b>	<b>8.5%</b> <b>14.2%</b>
Bayes SPM trong RGB (Bran và Mason 2000)	<b>93.4 %</b>	<b>19.8%</b>
Maximum Entropy Model trong RGB (Jedynak và al. 2002)	<b>80%</b>	<b>8%</b>
Gaussian Mixture models trong RGB (Jones và Rehg 1999)	<b>80%</b> <b>90%</b>	<b>~9.5%</b> <b>~15.5%</b>
SOM in TS (Brown và al. 2002)	<b>78%</b>	<b>32%</b>
Elliptical boundary model trong CIE – xy (Lee và Yoo 2002)	<b>90%</b>	<b>20.9%</b>
Single Gaussian trong Cb và Cr (Lee và Yoo 2002)	<b>90%</b>	<b>33.3%</b>
Gaussian Mixture trong IQ (Lee và Yoo 2002)	<b>90%</b>	<b>30,0%</b>
Thresholding của trục I trong YIQ (Brand và Mason 2000)	<b>94.7%</b>	<b>30.2%</b>

**Bảng 3-1: Kết quả nhận biết đúng và sai của các phương pháp**



### **3.1.6. Đánh giá phương pháp**

Ưu điểm chính của các phương pháp sử dụng các ngưỡng để phân lớp điểm ảnh là màu da hay không đó là tính đơn giản và tính trực giác cao trong các quy tắc phân lớp. Tuy nhiên, điểm khó khăn đó là cần phải tìm được cả một không gian màu tốt và các quy tắc xứng đáng trong không gian đó. Phương pháp được đề xuất hiện này sử dụng thuật toán máy học để có thể tìm được không gian và các quy tắc thích hợp, tuy nhiên đề xuất này vẫn đang là một vấn đề mở trong tương lai

Các phương pháp sử dụng mô hình hóa không tham số thật sự nhanh trong cả việc huấn luyện và phân lớp, độc lập với phân bố hình dạng của màu da và cả không gian màu. Tuy nhiên, phương pháp này lại yêu cầu quá nhiều bộ nhớ lưu trữ và phụ thuộc cố định vào tập dữ liệu huấn luyện.

Các phương pháp mô hình hóa có tham số cũng xử lý khá nhanh. Hơn nữa chúng lại có khả năng tự tạo ra các dữ liệu huấn luyện phù hợp, chúng được miêu tả bằng một số lượng không nhiều các tham số và đặc biệt chúng cần không đáng kể bộ nhớ lưu trữ. Tuy nhiên, chúng có thể sẽ thực sự chậm trong cả huấn luyện và làm việc, và hiệu năng của chúng phụ thuộc nhiều vào hình dạng của phân phối màu da. Bên cạnh đó, hầu hết các phương pháp mô hình hóa màu da có tham số đều bỏ qua những thống kê về màu không phải là tham số.

### **3.1.7. Chọn lựa không gian màu và phương pháp mô hình hóa dùng để nhận biết màu da cho đồ án.**

Với mục đích là sử dụng nhận biết màu da để tiến hành phân vùng màu da, giảm không gian tìm kiếm khuôn mặt trong ảnh màu. Vì vậy, phương pháp cần thiết cho đồ án phải có hiệu năng cao, thời gian thực hiện nhanh, yêu cầu bộ nhớ không lớn. Thời gian huấn luyện không phải là vấn đề. Không gian nhớ phải phù hợp với phương pháp nhận biết màu da. Chính vì vậy, xem xét tất cả các phương pháp, các đặc trưng cũng như hiệu năng của từng phương pháp. Chúng ta sẽ sử dụng phương pháp mô hình hóa có tham số để tiến hành phân vùng màu da cho ảnh. Cụ thể phương pháp được lựa chọn đó là phương pháp mô hình hóa màu da có tham số dựa trên phân phối Gaussian. Và không gian màu được lựa chọn đó là không gian RGB do đặc điểm đơn, sự phân biệt rõ ràng giữa độ sáng và các thành phần màu. Đồng thời, để nâng cao hiệu năng và

giảm thời gian tính toán cho phương pháp chọn lựa này, chúng ta sẽ xem xét các quy tắc được đề xuất trong nhóm phương pháp dùng ngưỡng để lọc khỏi tạo loại bỏ những điểm ảnh chắc chắn không phải màu da.

## **3.2. Nhận biết phân vùng màu da**

Với mỗi ảnh đầu vào ta tiến hành nhận biết và phân vùng màu da cho ảnh. Quá trình nhận biết là quá trình phân lớp các pixel trong ảnh ra làm hai lớp, lớp là pixel màu da và lớp pixel không phải màu da. Sau quá trình nhận biết, chúng ta sẽ tiến hành phân vùng để sử dụng làm giảm không gian tìm kiếm trong quá trình phát hiện mặt người. Nghĩa là chúng ta sẽ chỉ tìm kiếm mặt người trên vùng màu da được nhận biết.

### **3.2.1. Lọc khởi tạo**

Như đã từng đề cập, trước khi một pixel được nhận biết nó có là pixel màu da hay không nó sẽ được đưa qua một bộ lọc khởi tạo. Thực chất, bộ lọc này là một tập các quy tắc được chọn lọc từ các phương pháp chọn ngưỡng để phân vùng màu da. Tập quy tắc này được kiểm định từ thực nghiệm là đúng đắn với mọi màu da cần nhận biết.

Sau khi tham khảo một số các quy tắc được đề xuất, tiến hành chọn lọc và thử nghiệm trên nhiều ảnh màu khác nhau, đồ án đã đề xuất một tập các quy tắc sau nhằm giúp lọc ngay từ ban đầu các pixel không phải màu da. Điều này cho phép giảm khối lượng tính toán của ảnh đáng kể.

Tập quy tắc này được phát biểu trong không gian màu RGB như sau:

Một pixel sẽ không phải là pixel màu da nếu nó thỏa mãn một trong các điều kiện sau đây:

1.  $(B > 160 \ \&\& \ R < 180 \ \&\& \ G < 180)$  - *Quá nhiều thành phần xanh lơ (Blue)*

2.  $(G > 160 \ \&\& \ R < 180 \ \&\& \ B < 180)$  - *Quá nhiều thành phần xanh lá cây (Green)*

3.  $(B < 100 \ \&\& \ R < 100 \ \&\& \ G < 100)$  - *Quá tối*

4.  $(G > 200)$  – *pixel màu xanh lá cây Green*

5.  $(R < 102 \ \&\& \ G > 100 \ \&\& \ B > 110 \ \&\& \ G < 140 \ \&\& \ B < 160)$   
– *Màu xanh đại dương*

Hoặc là:

$$1. R > 95 \ \&\& \ G > 40 \ \&\& \ B < 20$$

$$2. \max\{R, G, B\} - \min\{R, G, B\} > 15$$

$$3. |R - G| > 15 \ \&\& \ R > G \ \&\& \ R > B$$

Hoặc là:

$$1. R > 220 \ \&\& \ G > 210 \ \&\& \ B > 170 \ \&\& \ |R - G| = 15 \ \&\& \ R > B \\ \&\& \ G > B$$

### 3.2.2. Nhận biết màu da

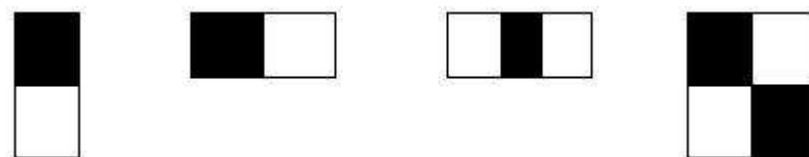
Kết quả của quá trình nhận biết có thể được thấy trong hình cạnh bên. Ảnh phân vùng sẽ có dạng ảnh nhị phân. Pixel màu da được giữ nguyên, còn pixel không phải màu da sẽ được đại diện bởi màu đen.



Hình 3-1: Ảnh trước và sau khi nhận biết màu da

### 3.3. Trích chọn đặc trưng Haar - like

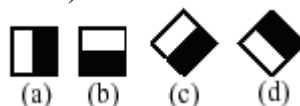
Đặc trưng Haar like do Viola và Jones công bố, gồm 4 đặc trưng cơ bản để xác định khuôn mặt người. Mỗi đặc trưng Haar-like là sự kết hợp của hai hay ba hình chữ nhật "trắng" hay "đen" như trong hình sau:



Hình 3-2: 4 đặc trưng Haar-like cơ bản

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

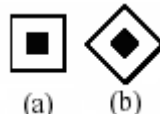
1. Đặc trưng cạnh (edge features):



2. Đặc trưng đường (line features):



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



Hình 3-3: Các đặc trưng mở rộng của các đặc trưng Haar-like cơ sở

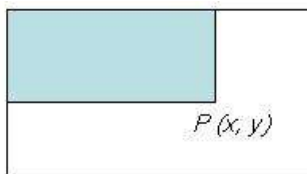
Dùng các đặc trưng trên, ta có thể tính được giá trị của đặc trưng Haar-like là sự chênh lệch giữa tổng của các pixel của các vùng đen và các vùng trắng như trong công thức sau :

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

Sử dụng giá trị này, so sánh với các giá trị của các giá trị pixel thô, các đặc trưng Haar-like có thể tăng/giảm sự thay đổi in-class/out-of-class (bên trong hay bên ngoài lớp khuôn mặt người), do đó sẽ làm cho bộ phân loại dễ hơn. Như vậy ta có thể thấy rằng, để tính các giá trị của đặc trưng Haar-like, ta phải tính tổng của các vùng pixel trên ảnh. Nhưng để tính toán các giá trị của các đặc trưng Haar-like cho tất cả các vị trí trên ảnh đòi hỏi chi phí tính toán khá lớn, không đáp ứng được cho các ứng dụng đòi hỏi tính run-time. Do đó Viola và Jones đưa ra một khái niệm gọi là Integral Image, là một mảng 2 chiều với kích thước bằng với kích của ảnh cần tính các đặc trưng Haar-like, với mỗi phần tử của mảng này được tính bằng cách tính tổng của điểm ảnh phía trên (dòng-1) và bên trái (cột-1) của nó. Bắt đầu từ vị trí trên, bên trái đến vị trí

dưới, phải của ảnh, việc tính toán này đơn thuần chỉ dựa trên phép cộng số nguyên đơn giản, do đó tốc độ thực hiện rất nhanh.

$$P(x,y) = \sum_{x' \leq x, y' \leq y} i(x', y') \quad (4.1)$$



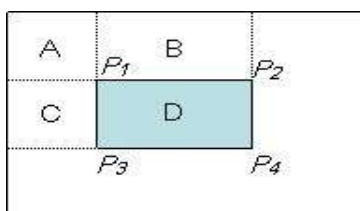
Hình 3-4: Cách tính Integral Image của ảnh

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

Giả sử ta cần tính tổng các giá trị mức xám của vùng D như trong hình 4, ta có thể tính như sau:

$$D = A + B + C + D - (A+B) - (A+C) + A \quad (4.2)$$

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



$$D = \underbrace{(x_4, y_4)}_{A + B + C + D} - \underbrace{(x_2, y_2)}_{(A+B)} - \underbrace{(x_3, y_3)}_{(A+C)} + \underbrace{(x_1, y_1)}_A \quad (4.3)$$

Hình 3-5: Ví dụ cách tính nhanh các giá trị mức xám của vùng D trên ảnh

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

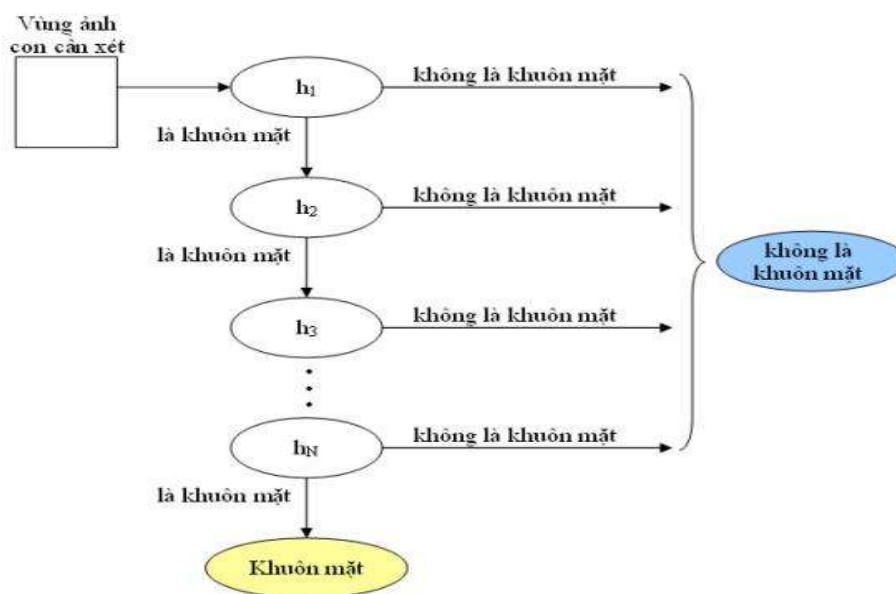
đoán một cách gần gũi một chút, còn bộ phân loại mạnh có thể đưa ra câu trả lời chính xác trên 60%.

### 3.4. Huấn luyện dò tìm khuôn mặt

AdaBoost là một bộ phân loại mạnh phi tuyến phức dựa trên hướng tiếp cận boosting được Freund và Schapire đưa ra vào năm 1995 [2]. Adaboost cũng hoạt động trên nguyên tắc kết hợp tuyến tính các weak classifiers để hình thành một strong classifier.

Là một cải tiến của tiếp cận boosting, AdaBoost sử dụng thêm khái niệm trọng số (weight) để đánh dấu các mẫu khó nhận dạng. Trong quá trình huấn luyện, cứ mỗi weak classifiers được xây dựng, thuật toán sẽ tiến hành cập nhật lại trọng số để chuẩn bị cho việc xây dựng weak classifier kế tiếp: tăng trọng số của các mẫu bị nhận dạng sai và giảm trọng số của các mẫu được nhận dạng đúng bởi weak classifier vừa xây dựng. Bằng cách này weak classifier sau có thể tập trung vào các mẫu mà các weak classifiers trước nó làm chưa tốt. Sau cùng, các weak classifiers sẽ được kết hợp tùy theo mức độ tốt của chúng để tạo nên strong classifier.

Viola và Jones dùng AdaBoost kết hợp các bộ phân loại yếu sử dụng các đặc trưng Haar-like theo mô hình phân tầng (cascade) như sau:



Hình 3-6: Mô hình phân tầng kết hợp các bộ phân loại yếu để xác định khuôn mặt

Trong đó,  $h_k$  là các bộ phân loại yếu, được biểu diễn như sau:

$$H_k = \begin{cases} 1 & \text{nếu } p_k f_k(x) < p_k \theta_k \\ 0 & \text{ngược lại} \end{cases}$$

$x$ : cửa sổ con cần xét

$\theta_k$ : ngưỡng ( $\theta = \text{teta}$ )

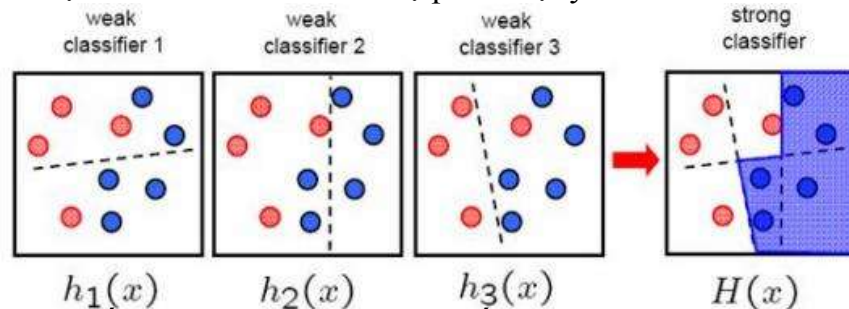
$f_k$ : giá trị của đặc trưng Haar-like

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

AdaBoost sẽ kết hợp các bộ phân loại yếu thành bộ phân loại mạnh như sau:

$$H(x) = \text{sign}(a_1 h_1(x) + a_2 h_2(x) + \dots + a_n h_n(x)) \quad (a = \text{alpha})$$

Với:  $a_t \geq 0$  là hệ số chuẩn hoá cho các bộ phân loại yếu



Hình 3-7: Kết hợp các bộ phân loại yếu thành bộ phân loại mạnh

### 3.5. Quá trình dò tìm khuôn mặt

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

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

- Duyệt toàn bộ các vị trí  $(x, y)$  cửa sổ với kích thước  $(w_0 \times h_0)$ , với mỗi vị trí tiến hành:
  - Áp dụng bộ dò tìm phân tầng để xác định có phải mặt người hay không
  - Nếu là mặt người tại vị trí  $(x, y)$  thì thực tế mặt người tại vị trí  $(x * \text{scale}^{\text{step}}, y * \text{scale}^{\text{step}})$  và kích thước cửa sổ là  $(w_0 * \text{scale}^{\text{step}}, h_0 * \text{scale}^{\text{step}})$
- Gán  $w_1 = w / \text{scale}$  và  $h_1 = h / \text{scale}$

- Thu nhỏ ảnh từ kích thước  $(w \times h)$  đến  $(w_1 \times h_1)$
- Gán  $w = w_1$  và  $h = h_1$
- $step = step + 1$

Nhận xét :

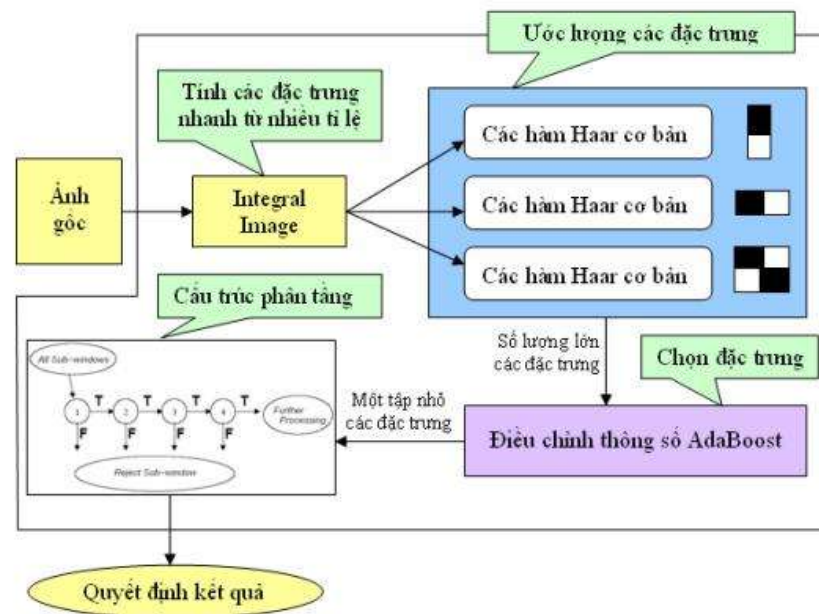
Hệ số co scale quyết định độ mịn của các cửa sổ dò tìm, nếu như scale càng nhỏ ( $\geq 1$ ) thì càng có nhiều cửa sổ dò tìm nên càng chính xác hơn.

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

- Bước 1 : kích thước từ  $(w_0, h_0)$  đến  $(w_0 \cdot scale, h_0 \cdot scale)$
- Bước 2 : kích thước từ  $(w_0 \cdot scale, h_0 \cdot scale)$  đến  $(w_0 \cdot scale^2, h_0 \cdot scale^2)$
- .....

Bước n : kích thước từ  $(w_0 \cdot scale^{n-1}, h_0 \cdot scale^{n-1})$  đến  $(w_0 \cdot scale, h_0 \cdot scale)$

### 3.6. Hệ thống xác định vị trí khuôn mặt người



Hình 3-8: Hệ thống xác định vị trí khuôn mặt người (Face detection system)

Như trong hình trên, từ ảnh gốc ban đầu, ta sẽ được tính Integral Image, là mảng 2 chiều với phần tử  $(x, y)$  sẽ được tính bằng tổng của các phần tử  $(x', y')$  với  $x' < x$  và  $y' < y$ , mục đích là để tính nhanh tổng của các giá trị mức xám của



một vùng hình chữ nhật bất kỳ trên ảnh gốc. Các vùng ảnh con này sẽ được đưa qua các hàm Haar cơ bản để ước lượng đặc trưng, kết quả ước lượng sẽ được đưa qua bộ điều chỉnh AdaBoost để loại bỏ nhanh các đặc trưng không có khả năng là đặc trưng của khuôn mặt người. Chỉ có một tập nhỏ các đặc trưng mà bộ điều chỉnh AdaBoost cho là có khả năng là đặc trưng của khuôn mặt người mới được chuyển sang cho bộ quyết định kết quả. Bộ quyết định sẽ tổng hợp kết quả là khuôn mặt người nếu kết quả của các bộ phân loại yếu trả về là khuôn mặt người.

## CHƯƠNG 4: CÀI ĐẶT ỨNG DỤNG

### 4.1. Môi trường TEST

Phần cứng : Một máy tính Pentum IV 2.8 Ghz.

Card đồ họa 128 MB.

Ram 512 MB .

Phần mềm : Ngôn ngữ sử dụng: VC++ trong bộ Visual Studio 2005.

Bộ cài OpenCV và OpenCV.NET

### 4.2. Một số giao diện chính



Hình 4-1: Giao diện chính của chương trình



Hình 4-2: Ảnh sau khi được mở



Hình 4-3: Ảnh sau khi được lọc màu da



Hình 4-4: Ảnh kết quả phát hiện khuôn mặt

### 4.3. Nhận xét

Chương trình cho kết quả tương đối tốt, dò tìm được hầu hết các khuôn mặt. Tuy nhiên vẫn còn những phát hiện thừa hoặc thiếu, điều này là do tập huấn luyện chưa được chính xác và các tư thế cũng như màu da của người chụp chưa được chuẩn xác như đứng lệch về một phía, màu da bị chói sáng quá, ...

## KẾT LUẬN

Phát hiện khuôn mặt là bài toán cơ bản và quan trọng trong lĩnh vực thị giác máy tính. Những công nghệ hiện có đang còn rất non trẻ và còn nhiều vấn đề cần phải tiếp tục nghiên cứu. Với đề tài “Dò tìm khuôn mặt trong ảnh dựa vào màu da”, khoá luận đã trình bày tổng quan một số phương pháp phát hiện khuôn mặt, tập trung trình bày hai phương pháp “phát hiện khuôn mặt dựa vào màu da” và “phát hiện khuôn mặt sử dụng đặc trưng Haar và AdaBoost”. Khoá luận đã xây dựng được một hệ thống phát hiện khuôn mặt sử dụng bộ lọc màu da cùng đặc trưng Haar và bộ lọc AdaBoost cho kết quả phát hiện được khuôn mặt nhanh chóng và chính xác. Việc cài đặt thử nghiệm thuật toán trên bằng ngôn ngữ MS VC++ 2005 đã được thực hiện, bước đầu cho kết quả tốt.

Trong tương lai em sẽ tiếp tục nghiên cứu thêm để có thể hoàn thiện hơn nữa về độ chính xác và thời gian dò tìm, và cố gắng để ứng dụng được chương trình vào một số hệ thống phục vụ các lĩnh vực: giám sát, theo dõi, an ninh trí tuệ nhân tạo...

Tuy nhiên do hạn chế về điều kiện và thời gian, khoá luận sẽ không thể tránh khỏi những thiếu sót. Kính mong được sự đóng góp ý kiến của thầy cô và các bạn, để em có thể hoàn thiện tốt hơn đề tài nghiên cứu của mình trong đợt khoá luận của em.

Em xin trân trọng cảm ơn!

## TÀI LIỆU THAM KHẢO

[1]. Phạm Thế Bảo, Nguyễn Thành Nhật, Cao Minh Thịnh, Trần Anh Tuấn, Phan Phúc Doãn - Tổng quan các phương pháp xác định khuôn mặt người

[2]. Nguyễn Tuấn Nghĩa (2005) – Sử dụng mô hình Entropy cực đại nhận dạng màu da trên ảnh màu – Luận văn tốt nghiệp Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội.

[3]. CBCL Face Database, CMU and MIT,

<http://www.ai.mit.edu/projects/cbcl.old/software-datasets/FaceData2.html>

[4]. Tài liệu về OpenCV