

# Tổng quan các phương pháp xác định khuôn mặt người

Phạm Thế Bảo, Nguyễn Thành Nhựt, Cao Minh Thịnh, Trần Anh Tuấn, Phan Phúc Doãn

## I. 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 thật sự và rất nhiều của con người.

### 1. Định nghĩa bài toán xác định khuôn mặt người

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, ... [105].

### 2. Ứng dụng của phương pháp xác định khuôn mặt người

Có nhiều ứng dụng đã được và đang thiết kế, tôi chỉ xin đưa ra một số loại ứng dụng sau:

- Hệ thống tương tác giữa người và máy: giúp những người bị tật hoặc khiếm khuyết có thể trao đổi. Những người dùng ngôn ngữ tay có thể giao tiếp với những người bình thường. Những người bị bại liệt thông qua một số ký hiệu nháy mắt có thể biểu lộ những gì họ muốn, .... Đó là các bài toán điều bộ của bàn tay (hand gesture), điệu bộ

khuôn mặt, ... [5, 6, 7, 32, 54, 95, 118, 130].

- Nhận dạng người A [29, 38, 46, 55, 56, 58, 60, 61] có phải là tội phạm truy nã hay không? Giúp cơ quan an ninh quản lý tốt con người. Công việc nhận dạng có thể ở trong môi trường bình thường cũng như trong bóng tối (sử dụng camera hồng ngoại).
- Hệ thống quan sát, theo dõi [35, 35, 106] và bảo vệ. Các hệ thống camera sẽ xác định đầu là con người và theo dõi con người đó xem họ có vi phạm gì không, ví dụ xâm phạm khu vực không được vào, ....
- Lưu trữ (rút tiền ATM, để biết ai rút tiền vào thời điểm đó), hiện nay có tình trạng những người bị người khác lấy mất thẻ ATM hay mất mã số PIN và những người ăn cắp này đi rút tiền, hoặc những người chủ thẻ đi rút tiền nhưng lại báo cho ngân hàng là mất thẻ và mất tiền. Các ngân hàng có nhu cầu khi có giao dịch tiền sẽ kiểm tra hay lưu trữ khuôn mặt người rút tiền để sau đó đối chứng và xử lý [66, 81, 98, 133].
- Thẻ căn cước, chứng minh nhân dân (Face Identification) [114].
- Điều khiển vào ra: văn phòng, công ty, trụ sở, máy tính, Palm, .... Kết hợp thêm vân tay và mống mắt. Cho phép nhân viên được ra vào nơi cần thiết, hay mỗi người sẽ đăng nhập máy tính cá nhân của mình mà không cần nhớ tên đăng nhập cũng như mật khẩu mà chỉ cần xác định thông qua khuôn mặt [44].
- An ninh sân bay, xuất nhập cảnh (hiện nay cơ quan xuất nhập cảnh Mỹ đã áp dụng).

Dùng để xác thực người xuất nhập cảnh và kiểm tra có phải là nhân vật khủng bố không.

- Tương lai sẽ phát triển các loại thẻ thông minh có tích hợp sẵn đặc trưng của người dùng trên đó, khi bất cứ người dùng khác dùng để truy cập hay xử lý tại các hệ thống sẽ được yêu cầu kiểm tra các đặc trưng khuôn mặt so với thẻ để biết nay có phải là chủ thẻ hay không.
- Tìm kiếm và tổ chức dữ liệu liên quan đến con người thông qua khuôn mặt người trên nhiều hệ cơ sở dữ liệu lưu trữ thật lớn, như internet, các hãng truyền hình, .... Ví dụ: tìm các đoạn video có tổng thống Bush phát biểu, tìm các phim có diễn viên Lý Liên Kiệt đóng, tìm các trận đá banh có Ronaldo đá, ... [50, 94, 134].
- Hiện nay có nhiều hướng tiếp cận để xác định một ảnh có phải là ảnh khóa thân hay không? Khuôn mặt người được xem như một yếu tố để xác định cho một hướng tiếp cận mà được dùng gần đây [271, 272].
- Ứng dụng trong video phone [10].
- Phân loại trong lưu trữ hình ảnh trong điện thoại di động. Thông qua bài toán xác định khuôn mặt người và trích đặc trưng, rồi dựa vào đặc trưng này để sắp xếp lưu trữ, giúp người sử dụng dễ dàng truy tìm khi cần thiết [69, 105].
- Kiểm tra trạng thái người lái xe có ngủ gật, mất tập trung hay không, và hỗ trợ thông báo khi cần thiết [109].
- Phân tích cảm xúc trên khuôn mặt [112].
- Trong lãnh vực thiết kế điều khiển robot [42, 43, 124, 151, 236].
- Hãng máy chụp hình Canon đã ứng dụng bài toán xác định khuôn mặt người vào máy chụp hình thể hệ mới để cho kết quả hình ảnh đẹp hơn, nhất là khuôn mặt người [277].

## II. PHƯƠNG PHÁP XÁC ĐỊNH KHUÔN MẶT NGƯỜI

Có nhiều nghiên cứu tìm phương pháp xác định khuôn mặt người, từ ảnh xám đến ngày nay là ảnh màu. Tôi sẽ trình bày một cách tổng quát nhất những hướng giải quyết chính cho bài toán, từ những hướng chính này nhiều tác giả thay đổi một số ý nhỏ bên trong để có kết quả mới.

Dựa vào tính chất của các phương pháp xác định khuôn mặt người trên ảnh. Các phương pháp này được chia làm bốn [9] hướng tiếp cận chính. Ngoài bốn hướng này, nhiều nghiên cứu có khi liên quan đến không những một hướng tiếp cận mà có liên quan nhiều hơn một hướng chính:

- **Hướng tiếp cận dựa trên tri thức:** Mã hóa các hiểu biết của con người về các loại khuôn mặt người thành các luật. Thông thường các luật mô tả quan hệ của các đặc trưng.
- **Hướng tiếp cận dựa trên đặc trưng không thay đổi:** Mục tiêu các thuật toán đi tìm các đặc trưng mô tả cấu trúc khuôn mặt người mà các đặc trưng này sẽ không thay đổi khi tư thế khuôn mặt, vị trí đặt thiết bị thu hình hoặc điều kiện ánh sáng thay đổi.
- **Hướng tiếp cận dựa trên so khớp mẫu:** Dùng các mẫu chuẩn của khuôn mặt người (các mẫu này được chọn lựa và lưu trữ) để mô tả cho khuôn mặt người hay các đặc trưng khuôn mặt (các mẫu này phải chọn làm sao cho tách biệt nhau theo tiêu chuẩn mà các tác giả định ra để so sánh). Các mối tương quan giữa dữ liệu ảnh đưa vào và các mẫu dùng để xác định khuôn mặt người.
- **Hướng tiếp cận dựa trên diện mạo:** Trái ngược hẳn với so khớp mẫu, các mô hình (hay các mẫu) được học từ một tập ảnh huấn luyện trước đó. Sau đó hệ thống (mô hình) sẽ xác định khuôn mặt người. Hay một số tác giả còn gọi hướng tiếp cận này là hướng tiếp cận theo phương pháp học.

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

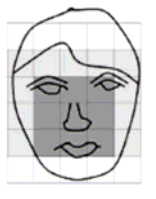
Trong hướng tiếp cận này, các luật sẽ phụ thuộc rất lớn vào tri thức của những tác giả nghiên cứu về bài toán xác định khuôn mặt người. Đây là hướng tiếp cận dạng top-down. Dễ dàng xây dựng các luật cơ bản để mô tả các đặc trưng của khuôn mặt và các quan hệ tương ứng. Ví dụ, một khuôn mặt thường có hai 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. Thường áp dụng quá trình xác định để giảm số lượng xác định sai.

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 mở rộng yêu cầu từ bài toán để xác định các khuôn mặt có nhiều tư thế khác nhau.



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



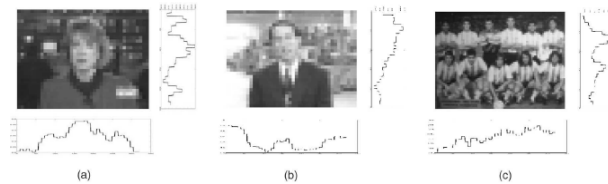
Hình 2: Một loại tri thức của người nghiên cứu phân tích trên khuôn mặt.

Yang và Huang [261] 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, hình 1. Các luật ở mức cao nhất để tìm ứng viên như: “vùng trung tâm khuôn mặt (phần tối hơn trong hình 2) có bốn phần với một mức độ đều cơ bản”, “phần xung quanh bên trên của một khuôn mặt (phần sáng hơn trong hình 2) có một mức độ đều cơ

bản”, và “mức độ khác nhau giữa các giá trị xám trung bình của phần trung tâm và phần bao bên trên là đáng kể”. Độ phân giải thấp nhất (mức mộ) 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 [200].

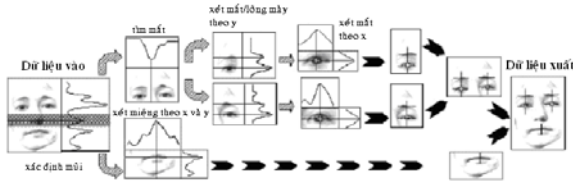
Kotopoulos và Pitas [200] đưa một phương pháp tương tự [191, 261] dùng trên độ phân giải thấp. Hai ông dùng phương pháp chiếu để xác định các đặc trưng khuôn mặt, Kanade đã thành công với phương pháp chiếu để xác định biên của khuôn mặt [191]. Với  $I(x,y)$  là giá trị xám của một điểm trong ảnh có kích 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:  $HI(x) = \sum_{y=1}^n I(x,y)$  và

$VI(y) = \sum_{x=1}^m I(x,y)$ . 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 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 3.b. Nếu ảnh có nhiều khuôn mặt thì sẽ không xác định được, hình 3.c.



Hình 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



Hình 4: Chiều từng phần ứng viên để xác định khuôn mặt.

Fan [82] phân đoạn ảnh màu để tìm cạnh thông qua thuật toán tăng vùng để xác định các ứng viên. Dùng đặc tính hình ellipse của khuôn mặt người để xác định ứng viên nào khuôn mặt người.

Kim [65] kết hợp thuật toán watershed cho các ảnh có nhiều độ phân giải cùng mô hình màu da người để tìm ứng viên, rồi xác định khuôn mặt người trong video. Tỷ lệ chính xác khoảng 87-94%. Phương pháp chỉ xử lý cho các frame ảnh chỉ có một khuôn mặt và ảnh này phải chụp thẳng chỉ có đầu và vai.

Sahbi và Boujemaa [8] sử dụng mạng neural học để ước lượng các tham số cho mô hình Gauss, mục đích để tìm ứng viên trên sắc màu da của người. Sau khi có ứng viên, hai ông chiếu lên hai trục: đứng và ngang để xác định khuôn mặt người.

Có nhiều nghiên cứu sau này sử dụng phương pháp chiếu để xác định khuôn mặt người. Min [80] dùng mô hình màu da không tham số, Baskan [76], Mateos [74], và Nicponski [45] xây dựng bộ lọc, để tìm ứng viên khuôn mặt, sau đó chiếu lên hai trục để xác định các thành phần khuôn mặt để xác định ứng viên đó có phải là khuôn mặt hay không. Còn Mateos và Chicote [34] 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 4. Tỷ lệ chính xác hơn 87%.

Farhad và Abdolhorsein [136] dùng tri thức về histogram để xác định khuôn mặt trong các frame liên tục trong một đoạn video. Tương tự, Hidekazu và Mamoru [100, 139] cũng dùng histogram, nhưng hai ông dùng thuật giải di truyền (Genetic Algorithm – GA) lại như là một phương pháp tìm kiếm ngẫu nhiên dựa vào đỉnh của biểu đồ màu của ảnh.

Rodrigues và Buf [132] dùng phương pháp chọn các keypoint trong nhiều tỷ lệ khác nhau, đặc biệt tác giả chỉ dùng các keypoint dư thừa dựa trên nhiều độ phân giải. Dựa trên quan hệ hình học của các thành phần khuôn mặt, hai ông nhóm các keypoint lại để xác định khuôn mặt người.

Fred [1140] dự trên tính chất đối xứng của khuôn mặt người, ông xem xét các phân bố trên histogram có tính chất gần đối xứng để xác định khuôn mặt người trong ảnh xám đơn có khuôn mặt chụp thẳng.

Berbar [279] 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.

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

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

### a) Các đặc trưng khuôn mặt

Sirohey đưa một phương pháp xác định khuôn mặt từ một ảnh có hình nền phức tạp [240]. Phương pháp dựa trên cạnh (dùng phương pháp Candy [155] 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 [157]. 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 [180]. 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 vấn đề, làm sao sử dụng các phép toán morphology và làm sao xác định khuôn mặt trên các vùng ứng viên.

Leung trình bày một mô hình xác suất để xác định khuôn mặt ở trong ảnh có hình nền phức tạp trên cơ sở một bộ xác định đặc trưng cục bộ và so khớp đồ thị ngẫu nhiên [205]. Ý chính là xem bài toán xác định khuôn mặt như là bài toán tìm kiếm với mục tiêu là tìm thứ tự các đặc trưng chắc chắn của khuôn mặt để tạo thành giống nhất một mẫu khuôn mặt. Dùng năm đặc trưng (hai mắt, hai lỗ mũi, phần nối giữa mũi và miệng) để mô tả một khuôn mặt. Luôn tính quan hệ khoảng cách với các đặc trưng cặp (như mắt trái, mắt phải), dùng phân bố Gauss để mô hình hóa. Một mẫu khuôn mặt được đưa ra thông qua trung bình tương ứng cho một tập đa hướng, đa tỷ lệ của bộ lọc đạo hàm Gauss. Từ một ảnh, các đặc trưng ứng viên được xác định bằng cách so khớp từng điểm ảnh khi lọc tương ứng với vector mẫu (tương tự mỗi tương quan), chọn hai ứng viên đặc trưng đứng đầu để tìm kiếm cho các đặc

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

Bên cạnh tính khoảng cách liên quan để mô tả quan hệ giữa các đặc trưng như Leung [154, 206]. Kendall [195] và [212] dùng lý thuyết xác suất thống kê về hình dáng. Dùng hàm mật độ xác suất (Probability Density Function - PDF) qua  $N$  điểm đặc trưng, tương ứng  $(x_i, y_i)$  là đặc trưng thứ  $i$  với giả sử dựa vào phân bố Gauss có  $2N$ -chiều. Các tác giả áp dụng phương thức cực đại khả năng (Maximum-Likelihood - ML) để xác định vị trí khuôn mặt. Một thuận lợi của phương pháp này là các khuôn mặt bị che khuất vẫn có thể xác định được. Nhưng phương pháp không xác định được đa khuôn mặt trong ảnh.

Yow và Cipolla [265, 266] trình bày một phương thức dựa vào đặc trưng, dùng số lượng lớn các dấu hiệu từ ảnh và cả dấu hiệu về ngữ cảnh. Đầu tiên dùng bộ lọc đạo hàm Gauss thứ hai, xác định các điểm mẫu chốt ở tại cực đại địa phương trong bộ lọc, rồi chỉ ra nơi có thể là đặc trưng. Giai đoạn hai, kiểm tra các cạnh xung quanh điểm mẫu chốt và nhóm chúng lại thành các vùng. Tiêu chuẩn để nhóm các cạnh là gần và tương tự hướng và cường độ. Đo lường các đặc tính vùng như: chiều dài cạnh, cường độ cạnh, và biến thiên cường độ được lưu trong một vector đặc trưng. Từ dữ liệu đặc trưng khuôn mặt đã được huấn luyện, sẽ tính được giá trị trung bình và ma trận hiệp phương sai của mỗi đặc trưng khuôn mặt. Một vùng là ứng viên khuôn mặt khi khoảng cách Mahalanobis giữa các vector đặc trưng đều dưới một ngưỡng. Rồi thông qua mạng Bayes để xác định ứng viên có phải là khuôn mặt không. Tỷ lệ chính xác là 85% [267], tuy nhiên mức độ sai là 28%, và chỉ hiệu quả với hình khuôn mặt có kích thước  $60 \times 60$  điểm ảnh. Phương pháp này được dùng thêm với mô hình đường viền linh hoạt [158, 267].

Takacs và Wechsler trình bày một phương pháp dựa trên tích đặc trưng vòng mạc và cử động theo dao động nhỏ của mắt [250]. Thuật toán hoạt động trên bản đồ hay vùng của các mẫu chốt, mô hình hóa lưới vòng mạc. Đầu tiên tính toán ước lượng thô vùng khuôn mặt trên cơ sở bộ lọc. Giai đoạn thứ hai

tính chế trên độ phân giải mịn hơn. Tỷ lệ sai là 4.69%.

Han phát triển một kỹ thuật trên cơ sở morphology để trích các đoạn giống mắt (eye-analogue) để xác định khuôn mặt người [182]. Ông nói rằng mắt và lông mày là đặc trưng nổi bật nhất và ổn định nhất của khuôn mặt con người, và nó rất hữu dụng để xác định khuôn mặt người. Ông định nghĩa các đoạn giống mắt như là các cạnh trên đường viền của mắt. Đầu tiên, các phép toán morphology như đóng, cắt bỏ sai khác, và phân ngưỡng để trích các điểm ảnh có giá trị cường độ thay đổi đáng kể. Các điểm ảnh này sẽ trở thành các điểm ảnh giống mắt. Sau đó một tiến trình gán nhãn để sinh các đoạn giống mắt. Các đoạn này được dùng để chỉ dẫn tìm kiếm các vùng tiềm năng có thể là khuôn mặt qua kết hợp các đặc tính hình học của mắt, mũi, lông mày, và miệng. Các vùng này sẽ được một mạng neural xem xét có phải là khuôn mặt không, giống [48]. Theo tác giả tỷ lệ chính xác là 94%.

Amit đưa ra phương thức xác định khuôn mặt dựa trên hình dáng và áp dụng cho các khuôn mặt chụp thẳng [145]. Có hai giai đoạn để xác định khuôn mặt người: tập trung và phân loại chi tiết. Làm có thứ tự các mảnh cạnh, các mảnh này được trích từ bộ xác định cạnh đơn giản thông qua sự khác biệt cường độ là quá trình tập trung. Khi có các ứng viên từ quá trình trên, dùng thuật toán CART [152] để xây dựng một cây phân loại từ các ảnh để huấn luyện, để xem xét ứng viên nào là khuôn mặt người.

Jin [90] dùng cấu trúc hình học của khuôn mặt người để tìm ứng viên khuôn mặt trong ảnh xám và hình nền không phức tạp. Mỗi ảnh chỉ có một khuôn mặt người, nhưng tư thế điều kiện ánh sáng, không cố định. Tỷ lệ chính xác khoảng 94.25% và thời gian khá nhanh.

Chan và Lewis [16] dùng kỹ thuật lọc để loại bớt tác động của ánh sáng, sau đó phân đoạn để tìm vị trí các ứng viên là con mắt. Từ các ứng viên này xây dựng mạng neural như Rowley [48] để xác định khuôn mặt người. Phương pháp này có thể xác định nhiều khuôn mặt trong một ảnh, các khuôn mặt này

có thể có các tư thế, vị trí, tỷ lệ khác nhau. Tỷ lệ chính xác là 53%.

Kruppa [21] dùng sắc màu của da người để tìm ứng viên, nhưng ông không xử lý cho từng điểm ảnh theo cách thông thường, mà ông dùng mô hình màu da người trên từng phần nhỏ rồi xử lý phân đoạn trên đó. Sau khi có ứng viên khuôn mặt, ông dùng một số đặc tính về hình dáng để xác định khuôn mặt người. Tỷ lệ chính xác là 85%.

Park dùng Gaze để tìm ứng viên góc mắt, miệng và tâm mắt [27]. Ông xây dựng SVM đã được học trước đó để xác định các vị trí ứng viên có phải là góc mắt, miệng, và tâm mắt hay không để theo vết con mắt người.

Sato [67] dùng quan hệ đường viền ở cằm của khuôn mặt. Tác giả chia làm hai trường hợp: thon dài và tròn để xem xét. Tác giả dùng GA để xem xét mối tương quan của đường cong, hình dáng khuôn mặt để xác định khuôn mặt.

Chai và Ngan [708] xây dựng phương pháp xác định khuôn mặt người dựa trên đặc trưng về: quan hệ hình học, mật độ, độ chói trong ảnh màu chỉ có đầu và vai của ứng viên để xác định. Kim [47] cũng phân đoạn để tìm ứng viên khuôn mặt, nhưng xác thực khuôn mặt thông qua các cấu trúc các đặc trưng mắt, mũi, miệng, và đường viền của ứng viên.

Jang [53] dùng phân bố màu da để phân đoạn tìm ứng viên rồi dùng các đặc trưng hình học để xác định khuôn mặt.

Christian và Jonh [135] xây dựng một loại đặc trưng mới, đó là đặc trưng về độ cong của các đường trên khuôn mặt để giải quyết vấn đề điều kiện ánh sáng. Từ đặc trưng cong này, hai ông quay lại phương pháp PCA để xác định khuôn mặt.

Juan và Narciso [111] xây dựng một không gian màu mới  $YC_g-C_r$  để lọc các vùng là ứng viên khuôn mặt dựa trên sắc thái của màu da người. Sau khi có ứng viên, hai ông dùng các quan hệ về hình dáng khuôn mặt, mức độ cân đối của các thành phần khuôn mặt để xác định khuôn mặt người. Tương tự, Chang và Hwang [127] cũng dùng một phương thức như [111], tỷ lệ chính xác hơn 80% trong ảnh xám.

Dae và Nam [116] xem xét các đặc trưng không thay đổi khi thay đổi tư thế của khuôn mặt bằng cách xem xét các quan hệ hình học. Sau đó ước

lượng các tư thế của khuôn mặt rồi xây dựng dữ liệu để xác định thông qua PCA. Tỷ lệ chính xác là 76%.

Jin [128] xây dựng một bộ lọc để xác định ứng viên khuôn mặt người theo màu da người. Từ ứng viên này tác giả xác định khuôn mặt người theo hình dáng khuôn mặt và các quan hệ đặc trưng về thành phần khuôn mặt, với mắt phải được chọn làm gốc tọa độ để xét quan hệ. Tỷ lệ chính xác cho khuôn mặt chụp thẳng trên 80%.

#### *b) 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 [147], gọi là kết cấu giống khuôn mặt (face-like texture). Tính kết cấu qua các đặc trưng thống kê thứ tự thứ hai (SGLD) [183] trên vùng có kích thước 16x16 điểm ảnh. Có ba loại đặc trưng được xem xét: màu da, tóc, và những thứ khác. Hai ông dùng mạng neural về mối tương quan cascade [170] cho phân loại có giám sát các kết cấu và một ánh xạ đặc trưng tự tổ chức Kohonen [199] 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 [165]. 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 [12] dùng phân bố màu da và thuật toán DoG (a Difference of Gauss) để tìm các ứng viên, rồi xác thực bằng một hệ thống học kết cấu của khuôn mặt.

Manian và Ross [88] 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%.

#### *c) Sắc màu của da*

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

#### *d) Đa đặc trưng*

Gần đây có nhiều nghiên cứu sử dụng các đặc trưng toàn cục như: màu da người, kích thước, và hình dáng để tìm các ứng viên khuôn mặt, rồi sau đó sẽ xác định ứng viên nào là khuôn mặt thông qua dùng các đặc trưng cục bộ (chi tiết) như: mắt, lông mày, mũi, miệng, và tóc. Tùy mỗi tác giả sẽ sử dụng tập đặc trưng khác nhau [70, 186].

Yachida đưa ra một phương pháp xác định khuôn mặt người trong ảnh màu bằng lý thuyết logic mờ [156, 259, 260]. 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 [241]. 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) [251, 252], các tiền ứng viên. 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 neural để biết vùng ứng viên nào là khuôn mặt người và vùng nào không phải khuôn mặt người. Tỷ lệ chính xác là 85%.

Dựa vào mức độ cân xứng của các mẫu khuôn mặt người để xác định khuôn mặt người [154]. Một bộ phân loại màu da/không phải màu da dùng trong không gian màu YES cho phép làm mịn các vùng kề có đường cong không mịn, sau khi lọc các vùng có thể là màu da người. Một mẫu khuôn mặt dạng ellipse được dùng để xem xét mức độ tương tự của các vùng có cùng màu da người với mẫu này thông qua khoảng cách Hausdorff [188]. Sau cùng, xác định tâm mắt thông qua các hàm tính giá trị dựa trên quan hệ cân đối của khuôn mặt và vị trí hai mắt. Đỉnh của mũi và tâm của miệng được ước lượng qua khoảng cách tâm mắt. Mặt hạn chế của phương pháp này là chỉ xác định trên ảnh chụp thẳng khuôn mặt, chỉ có duy nhất một khuôn mặt trong ảnh, và xác định được vị trí của cả hai mắt. Cũng có tác giả dùng phương pháp tương tự để giải quyết [245].

Trái ngược với phương pháp xử lý trên điểm ảnh, một phương pháp được xây dựng trên cấu trúc, màu sắc, và liên quan hình học đã được đề nghị [262]. Đầu tiên dùng phân đoạn đa tỷ lệ [144] để trích các vùng đồng đều trong ảnh dựa vào mô hình màu da người theo Gauss để có được các vùng có màu cùng với màu da người, gom các vùng này vào trong các vùng có hình dạng ellipse. Một vùng có hình dạng ellipse được xác định là một khuôn mặt người nếu tồn tại mắt miệng trong vùng đó. Tác giả cho biết có thể xác định các khuôn mặt ở các hướng khác nhau khi có thêm các đặc trưng phụ như: râu, mắt kính.

Kauth trình bày một biểu diễn dạng blob để trích đặc trưng, mà đặc trưng này dùng mô tả có ý nghĩa cấu trúc của đa phổ của ảnh chụp từ vệ tinh [194]. Mỗi vector đặc trưng tại một điểm ảnh bao gồm các tọa độ của điểm ảnh và liên quan theo các thành phần phổ (hay các thành phần kết cấu). Các điểm ảnh này được gom nhóm bằng cách dùng vector đặc trưng để có các vùng dính liền nhau, hoặc có dạng

blob. Mỗi vector đặc trưng bao gồm tọa độ ảnh và sắc màu được chuẩn hóa,

$$X = (x, y, \frac{r}{r+g+b}, \frac{g}{r+g+b}) \quad [218, 243].$$

Dùng một thuật toán tạo các vùng liên kết lại với nhau để tăng kích thước của blob và xem xét nếu ứng viên dạng blob nào thỏa mãn hình dáng kích thước khuôn mặt thì xem đó là khuôn mặt.

Phạm vi và màu sắc được Kim [197] dùng để xác định khuôn mặt người. Tính biểu đồ chênh lệch rồi phân đoạn dựa trên biểu đồ histogram với giả thuyết các điểm ảnh là nền sẽ có cùng độ sâu và số lượng sẽ nhiều hơn các điểm ảnh trong đối tượng. Dùng phân bố Gauss trong không gian màu RGB đã được chuẩn hóa, được các ứng viên rồi dùng phân loại để xác định cuối cùng ứng viên nào là khuôn mặt người. Cùng các tiếp cận này có Darrell [84].

Hsu được xem là người khá thành công khi xác định khuôn mặt người trong ảnh màu [1, 96]. Ông xây dựng một bộ phân loại để xác định các vị trí của ứng viên mắt và miệng dựa trên sắc màu đặc trưng của mắt và miệng. Trên quan hệ về khoảng cách của hai mắt và miệng để xác định ứng viên nào sẽ là khuôn mặt thông qua biến đổi Hough để có ứng viên nào gần giống dạng ellipse nhất.

Jesorsky [270] xác định cạnh của các đối tượng trong ảnh rồi so sánh hình dáng kết hợp dùng khoảng cách Hausdorff để đo mức độ tương tự của khuôn mặt người với các mẫu. Sau đó Kirchberg [17] cải tiến dùng mô hình Gen (Genetic Model) để phát sinh mô hình khuôn mặt người từ dữ liệu lộn xộn sau khi phân đoạn trong ảnh xám kết hợp khoảng cách Hausdorff. Mức độ chính xác khoảng 85%.

Yen và Nithianandan [66] dùng GA để trích các đặc trưng khuôn mặt, như mắt (lông mày), mũi, và miệng. Áp dụng hình thái khuôn mặt giống hình ellipse để xác định khuôn mặt bằng GA trong ảnh màu. Phương pháp này cho phép giải quyết trong điều kiện ánh sáng khác nhau, tư thế khuôn mặt khác nhau.

Chang [89] xem xét tính đa dạng về mặt của khuôn mặt người. Từ đây ông xây dựng mạng wavelet tích cực (Active Wavelet Network) để trích các đặc trưng của khuôn mặt rồi dùng hai phương



pháp làm giảm số chiều của không gian đặc trưng là LLE (Locally Linear Embedding) và LE (Lipschitz Embedding) và học cấu trúc đa dạng này để xác định khuôn mặt.

Daidi và Irek [117] trích các đặc trưng của khuôn mặt bằng sơ đồ phân bố tham số để xác định khuôn mặt người. Tỷ lệ chính xác cho ảnh xám và khuôn mặt được chụp thẳng là 91.4%.

Ehsan và Jonh [125] dùng tập hệ số Gabor wavelet ở các hướng khác nhau để trích các đặc trưng của khuôn mặt. Sau đó dùng entropy cục bộ để xác định khuôn mặt trong ảnh xám và khuôn mặt được chụp thẳng hay tựa thẳng nhưng có các vị trí khác nhau. Tỷ lệ chính xác là 94%.

Bao [281, 282] dùng sắc thái màu da người để xác định ứng viên trong ảnh màu. Tác giả đã xây dựng các luật mờ dựa vào hai loại đặc trưng: (1) bên ngoài và (2) bên trong. Đặc trưng bên ngoài gồm: tỷ lệ chiều cao, diện tích, chu vi, mức độ tròn, ... Đặc trưng bên trong gồm: quan hệ mức độ cân đối của hai mắt và miệng cũng như tỷ lệ khoảng cách với khuôn mặt. Phương pháp này cho phép xác định khuôn mặt ở nhiều tư thế, vị trí, mức độ nghiêng khác nhau trong môi trường phức tạp. Đặc biệt, tác giả đã xây dựng bộ điều khiển mờ để tách các khuôn mặt dính lẫn nhau. Tỷ lệ chính xác khoảng 87%-89%.

### 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.

Oh [119] phân đoạn để tìm ứng viên khuôn mặt, tác giả dùng các mẫu mặt có trước để so khớp với các vùng quan tâm để tìm vị trí mắt trong ứng viên.

Sau đó tiếp tục tìm miệng và lông mày để xác định ứng viên này có phải là khuôn mặt người hay không.

#### a) *Xác định các mẫu trước*

Sakai đã cố gắng thử xác định khuôn mặt người chụp thẳng trong ảnh [232]. 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) [163]. 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 [164].

Govindaraju đề nghị một phương thức xác định khuôn mặt người có hai giai đoạn để phát sinh các giả thuyết khuôn mặt và kiểm tra nó [177, 178, 179]. Một mô hình khuôn mặt được xây dựng trong các giai đoạn xác định đặc trưng bằng các cạnh. Các đặc trưng được mô tả như các đường cong của phía bên trái, đường viền tóc, phía bên phải của khuôn mặt được chụp thẳng. Dùng phép toán Marr-Hildreth để xác định cạnh. Sau đó dùng một bộ lọc để loại bỏ các đối tượng không tham gia vào xây dựng khuôn mặt. Liên kết các cặp của các đoạn đường viền trên cơ sở mức độ kề và các hướng liên quan. Xác định các góc để phân đoạn đường viền thành các đường cong đặc trưng. Gán nhãn các đường cong đặc trưng bằng cách kiểm tra thuộc tính hình học và các vị trí

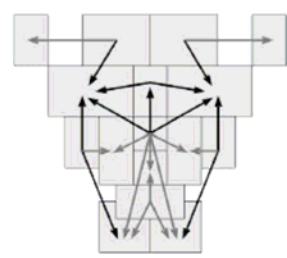
liên quan trong láng giềng của nó. Nối các cặp của các đường cong đặc trưng thông qua các cạnh nếu các thuộc tính của nó tương thích. So sánh các tỷ lệ của các cặp thuộc tính cho một cạnh và ấn định một giá trị tương ứng. Nếu giá trị của một nhóm của ba đường cong đặc trưng (với các nhãn khác nhau) thấp thì nhóm này sẽ trở thành một giả thuyết. Khi xác định khuôn mặt trong các bài báo thì thông tin phụ sẽ được dùng thêm đó là số lượng người trong ảnh để chọn giả thuyết tối ưu [178]. Tỷ lệ chính xác của phương pháp này là 70%, tuy nhiên các khuôn mặt phải được chụp thẳng và không bị che khuất. Venkatraman và Govindaraju dùng cách tiếp cận tương tự, nhưng dùng wavelet để trích cạnh [257].

Tsukamoto trình bày một mô hình hiệu quả khi dùng mẫu khuôn mặt (QMF) [253, 254]. Trong QMF, mỗi ảnh mẫu được chia thành nhiều khối, các đặc trưng hiệu quả được ước lượng cho mỗi khối. Tham số hóa một mẫu khuôn mặt theo: lightness và edgeness là các đặc trưng trong mô hình. Sau đó dùng các mẫu (đã được chia thành các khối) để tính giá trị faceness (mức độ là khuôn mặt) ở tại mọi vị trí của ảnh. Một khuôn mặt được xác định khi giá trị faceness vượt một ngưỡng đã được cho trước.

Hình chiếu được dùng như các mẫu để xác định khuôn mặt người [233]. Dùng PCA (phân tích thành phần chính – Principal Component Analysis - PCA) để có một tập hình chiếu cơ bản từ các mẫu khuôn mặt, hình chiếu được mô tả như một mảng các bit. Dùng đặc trưng hình chiếu riêng kết hợp biến đổi Hough để xác định khuôn mặt người. Sau đó một phương pháp xác định dựa trên đa loại mẫu để xác định các thành phần của khuôn mặt được trình bày [244]. Phương pháp này định nghĩa một số giả thuyết để mô tả các khả năng của các đặc trưng khuôn mặt. Với một khuôn mặt sẽ có một tập giả thuyết, lý thuyết DempsterShafer [166]. Dùng một nhân tố tin cậy để kiểm tra sự tồn tại hay không của các đặc trưng của khuôn mặt, và kết hợp nhân tố tin cậy này với một độ đo để xem xét có hay không có khuôn mặt trong ảnh.

Sinha dùng một tập nhỏ các bất biến ảnh trong không gian ảnh để mô tả không gian các mẫu ảnh [238, 239]. 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, 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 [222]. Ý tưởng của Sinha còn được áp dụng cho hệ thống thị giác của robot [151, 236]. Hình 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.



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

Phương pháp so khớp mẫu theo thứ tự để xác định khuôn mặt người do Miao trình bày [214]. 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, 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

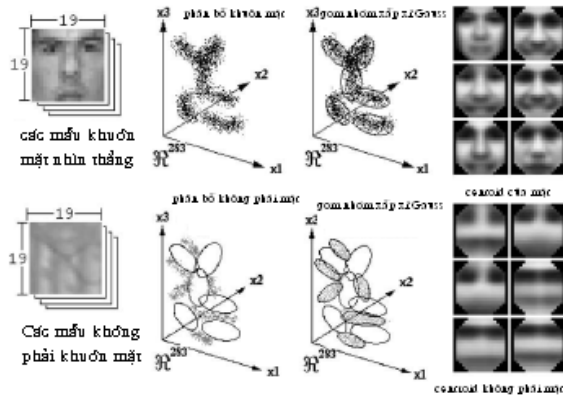
quả không tốt bằng xác định một khuôn mặt (chụp thẳng hoặc xoay) trong ảnh xám.

Wei và Lai [78] 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 [84] 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  $10^\circ$  [86].

Holst xây dựng một hệ thống từ các mẫu với các đặc trưng kép [92]: (1) thành phần, gồm: mắt, mũi, và miệng; (2) hình dáng khuôn mặt, trên độ phân giải thấp. Ông dùng hai phương pháp tìm kiếm trong không gian dữ liệu của mình để xác định khuôn mặt người.



Hình 6: Phân nhóm dữ liệu khuôn mặt và nhóm dữ liệu không phải khuôn mặt.

Proba và Zink lọc cạnh ở độ phân giải thấp rồi dùng biến đổi Hough để so khớp mẫu theo hướng cạnh để xác định hình dáng khuôn mặt ở dạng chụp hình thẳng ở dạng xám. Tỷ lệ chính xác trên 91% [25].

Shu và Jain xây dựng ngữ nghĩa khuôn mặt [85]. Ngữ nghĩa theo hình dáng và vị trí các thành phần khuôn mặt. Hai ông từ bộ ngữ nghĩa này xây dựng một đồ thị quan hệ để dễ dàng so khớp khi xác định khuôn mặt người.

Iwata [39] xây dựng mẫu mỗi đặc trưng gồm bốn đặc trưng theo bốn hướng: ngang, bên phải phía trên, đứng, và bên trái phía trên của khuôn mặt chụp thẳng hoặc gần thẳng trong ảnh xám. Để so khớp từng phần của mẫu kết hợp xác suất các láng giềng. Tỷ lệ chính xác của phương pháp này là gần 99%.

Keren [33] xây dựng khái niệm Antifaces để xác định khuôn mặt người (tổng quát cho các đối tượng 3-chiều). Dựa trên nhiều loại mẫu kết hợp giả thuyết phân bố xác suất để tìm những đối tượng không có mối tương quan để tìm khuôn mặt người. Ông cho biết, phương pháp này nhanh hơn eigenface và SVM và mức độ chính xác gần tương đương.

Feris [59] dùng mạng wavelet thứ nhất để xác định ứng viên khuôn mặt khi so khớp với các mẫu đã học trước. Sau đó tác giả dùng mạng wavelet thứ hai để xác định các thành phần như mắt, mũi, và miệng thông qua các đặc trưng góc cạnh. Từ các thành phần này xem xét tính hòa hợp để có quyết định cuối cùng ứng viên nào là khuôn mặt người.

#### b) 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 [268]. 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.

Một hướng tiếp cận dựa trên đường gấp khúc (snake) [193, 208] và các mẫu để xác định khuôn mặt [202]. Đầu tiên một ảnh sẽ được làm xoắn lại bởi một lọc làm mờ rồi dùng phép toán morphology để làm nổi bật cạnh lên. Dùng một đường gấp khúc có  $n$  điểm ảnh (giá trị  $n$  nhỏ) để tìm và ước lượng các đoạn cong nhỏ. Mỗi khuôn mặt được xấp xỉ bằng một ellipse và biến đổi Hough, rồi tìm một ellipse nổi trội nhất. Một tập có bốn tham số mô tả một ellipse được dùng như ứng viên để xác định

khuôn mặt. Với mỗi ứng viên, một phương thức tương tự như phương thức mẫu biến dạng [268] dùng để xác định các đặc trưng ở mức chi tiết. Nếu tìm thấy số lượng đáng kể các đặc trưng khuôn mặt và thỏa tỷ lệ cân đối thì xem như đã xác định được một khuôn mặt. Lam và Yan cũng dùng đường gấp khúc để xác định vị trí đầu với thuật toán greedy để cực tiểu hóa hàm năng lượng [203].

Thay vì dùng đường gấp khúc thì Huang và Su [13] dùng lý thuyết dòng chảy để xác định đường viền khuôn mặt dựa trên đặc tính hình học. Hai ông dùng lý thuyết tập đồng mức (Level Set) để loang từ các khởi động ban đầu để có được các khuôn mặt người.

Lanitis mô tả một phương pháp biểu diễn khuôn mặt người với cả hai thông tin: hình dáng và cường độ [204]. 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. Dùng tiếp cận như Kirby và Sirovich [198] để mô tả cường độ bề ngoài của hình dáng đã được chuẩn hóa. Một PDM có hình dáng như khuôn mặt dùng để xác định khuôn mặt bằng mô hình hình dáng tích cực (Active Shape Model - ASM) để tìm kiếm và ước lượng vị trí khuôn mặt cũng như các tham số về hình dáng. Các mảnh của khuôn mặt được làm biến dạng về hình dáng trung bình rồi trích các tham số cường độ. Các tham số hình dáng và cường độ được dùng để phân loại. Cootes và Taylor áp dụng cách tiếp cận này để xác định khuôn mặt [161]. Đầu tiên, hai ông định nghĩa một vùng hình chữ nhật chứa các đặc trưng quan tâm. Dùng phân tích nhân tố [146] làm vừa các đặc trưng huấn luyện để có một hàm phân bố. Có được các đặc trưng là ứng viên nếu độ đo xác suất vượt qua một ngưỡng khi dùng ASM. Sau khi huấn luyện xong có thể xác định khuôn mặt người. Hướng tiếp cận theo ASM được mở rộng bằng hai loại Kalman để ước lượng các tham số về hình dáng và cường độ dùng để theo vết khuôn mặt người [169].

#### 4. Hướng tiếp cận dựa trên diện mạo

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

Có nhiều phương pháp áp dụng xác suất thống kê để 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  $p(x | \text{khoảng phaikhuôn mặt})$ . 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  $p(x | \text{khoảng phaikhuôn mặt})$  là đa phương thức, và chưa thể hiểu nếu xây dựng các dạng tham số hóa một cách tự nhiên cho  $p(x | \text{khuôn mặt})$  và  $p(x | \text{khoảng phaikhuôn mặt})$ . 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  $p(x | \text{khoảng phaikhuôn mặt})$ .

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 [255], hoặc xây dựng mặt quyết định phi tuyến bằng mạng neural đa tầng [48]. 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 [220].

#### a) *Eigenface*

Kohonen đã đưa ra phương pháp dùng vector riêng để nhận dạng khuôn mặt [199], ông dùng một mạng neural đơn giản để chứng tỏ khả năng của phương pháp này trên các ảnh đã được chuẩn hóa. Mạng neural 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 [198]. Tính chất này dựa trên biến đổi Karhunen-Löve [176, 192, 211], mà còn được gọi dưới một cái tên khác là PCA [189] và biến đổi Hotelling [104]. Ý tưởng này được xem là của Pearson trình bày đầu tiên vào năm 1901 [223] và sau đó là Hotelling vào năm 1933 [185]. 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).

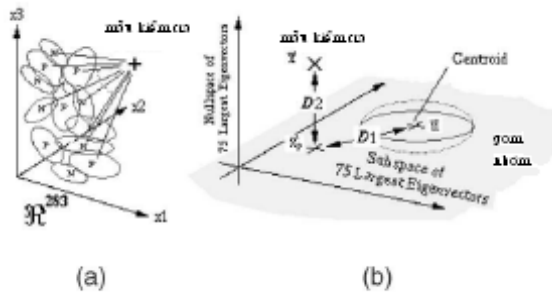
Turk và Pentland áp dụng PCA để xác định và nhận dạng khuôn mặt [255]. Tương tự [198], dùng PCA trên tập huấn luyện ảnh các khuôn mặt để sinh các ảnh riêng (còn gọi là eigenface) để tìm một không gian con (không gian khuôn mặt) trong không gian ảnh. Các ảnh khuôn mặt được chiếu vào không gian con này 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 [23] phát triển cho ảnh màu, bằng cách phân đoạn ảnh để tìm ứng đề không gian tìm kiếm bớt đi.

#### b) *Hướng tiếp cận dựa trên phân bố*

Sung và Poggio phát triển một hệ thống xác định khuôn mặt người dựa trên phân bố [246, 247], chứng tỏ bằng cách dùng phân bố các các mẫu ảnh cùng một lớp đối tượng có thể được học từ các mẫu negative và positive. Hệ thống của hai ông bao gồm hai thành phần: mô hình phân bố của các mẫu là khuôn mặt/không phải khuôn mặt và một phân loại đa tầng dựa vào thị giác. Mỗi mẫu là khuôn mặt và không phải là khuôn mặt được chuẩn hóa và xử lý thành ảnh có kích thước  $19 \times 19$  điểm ảnh và xem như một vector hay mẫu có 361-chiều. Sau đó các mẫu được nhóm vào các nhóm, mỗi nhóm gồm sáu mẫu cùng loại là khuôn mặt hoặc nhóm không phải là khuôn mặt bằng thuật toán k-trung bình (k-mean), hình 6. Mỗi nhóm sẽ được mô tả như một hàm Gauss đa chiều với một ảnh trung bình và ma trận hiệp phương sai. Hình 7 cho thấy cách tính khoảng cách của hai ông. Hai độ đo khoảng cách dùng để tính khoảng cách giữa ảnh đưa vào và tâm của nhóm. Thành phần khoảng cách đầu tiên là khoảng cách Mahalanobis được chuẩn hóa giữa hình chiếu của mẫu cần kiểm tra và tâm của nhóm, tính trong không gian con có số chiều thấp hơn, được mô tả bằng 75 vector riêng lớn nhất. Thành phần khoảng cách thứ hai là khoảng cách Euclide giữa mẫu cần kiểm tra và hình chiếu của nó trong không gian con có 75- chiều này. Dùng hai khoảng cách này để xác định khoảng cách từ mẫu cần kiểm tra đến tâm một nhóm. Từ nay chúng ta có thể biết mẫu cần kiểm tra gần nhóm nào nhất. Bước cuối cùng dùng mạng đa tầng (Multilayer Perceptron Network – MLP) để

phân loại dựa vào 12 cặp khoảng cách (có 12 nhóm) khi mạng này đã được huấn luyện trước đó. Dễ dàng chọn mẫu khuôn mặt để huấn luyện, nhưng không dễ để chọn mẫu không phải là khuôn mặt để huấn luyện. Dùng phương pháp bootstrap để giải quyết vấn đề này. Bắt đầu từ tập nhỏ không phải khuôn mặt trong tập mẫu huấn luyện huấn luyện MLP. Dùng bộ xác định khuôn mặt người để xác định mặt người trên một dãy các ảnh ngẫu nhiên, sau đó chọn các mẫu không phải khuôn mặt người mà bị xác định là khuôn mặt người xem như là mẫu không phải khuôn mặt người mới để huấn luyện tiếp tục. Phương pháp này bỏ qua vấn đề chọn mẫu nào trong các mẫu để tăng tính hiệu quả, có nhiều nghiên cứu sau này về vấn đề này [48, 220].

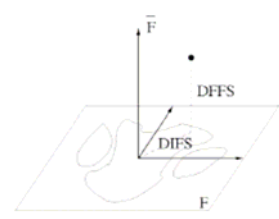


Hình 7: (a) Khoảng cách giữa mẫu cần kiểm tra và các nhóm;  
(b) hai thành phần khoảng cách.

Moghaddam và Pentland đưa ra một mô hình học theo xác suất dựa trên ước lượng mật độ trong không gian có số chiều cao bằng không gian riêng [216]. Hai ông dùng PCA để tìm không gian con để mô tả tốt nhất một tập các mẫu khuôn mặt người. Phương pháp này vẫn giữ lại các mối tương quan tuyến tính chính trong dữ liệu và loại bỏ các thứ yếu khác. Phương pháp này phân rã một không gian vector thành hai không gian con mà hai không gian con này loại trừ lẫn nhau và cũng bổ sung cho nhau: không gian con chính (không gian đặc trưng) và phần bù trực giao. Vì thế, mục tiêu mật độ được phân rã làm hai thành phần: mật độ trong không gian chính (dựa vào các thành phần chính) và phần bù trực giao, hình 8. Xây dựng hệ thống học dựa vào Gauss nhiều biến và Gauss hỗn hợp, hệ thống này học dựa trên thống kê các đặc trưng cục bộ của một khuôn mặt. Dùng các mật độ xác suất để xác định khuôn mặt trên cơ sở ước lượng khả năng cực đại. Phương pháp

này được áp dụng cho xác định khuôn mặt, mã hóa khuôn mặt, và nhận dạng. So sánh với hướng tiếp eigenface cổ điển [255], phương pháp này cho thấy hiệu quả hơn trong xác định và nhận dạng khuôn mặt [196].

Yang sử dụng một hỗn hợp nhiều phân tích hệ số làm tiêu chí xác định khuôn mặt. Phân tích hệ số (Factor Analysis – FA) là một phương pháp thống kê để mô hình hóa tính hiệp biến cấu trúc của dữ liệu có số chiều cao bằng cách dùng mô hình nhỏ các biến tiềm tàng. FA cũng tương tự PCA trong vài khía cạnh. Tuy nhiên, PCA không giống FA, không định nghĩa một mô hình mật độ thích hợp cho dữ liệu. Hơn nữa, PCA không hiệu quả khi có nhiều độc lập trong các đặc trưng của dữ liệu. Tổng hợp từ [148, 150, 167, 168] cho thấy các mẫu được chiếu từ các lớp khác nhau vào không gian con PCA thường có thể không hiệu quả. Trong các trường hợp khi các mẫu có một cấu trúc chắc chắn, dùng PCA sẽ cho kết quả khá tốt. Hinton dùng FA để nhận dạng các con số, ông đã so sánh FA và PCA [184]. Một mô hình hỗn hợp của các phân tích hệ số được mở rộng để nhận dạng khuôn mặt người [174]. Cả hai nghiên cứu đều cho thấy FA tốt hơn PCA. Từ tư thế, hướng, cảm xúc, và ảnh hưởng ánh sáng trên diện mạo của khuôn mặt người, phân bố các khuôn mặt trong không gian ảnh có thể được biểu diễn tốt hơn bằng một mô hình mật độ đa phương thức khi mỗi phương thức giữ các đặc tính chắc chắn của diện mạo chắc chắn của khuôn mặt. Họ đã trình bày một mô hình theo xác suất khi dùng một hỗn hợp các phân tích hệ số (Mixture of Factor Analyzer – MFA) để xác định khuôn mặt người. Dùng thuật toán EM để ước lượng các tham số trong mô hình hỗn hợp.



Hình 8: Phân rã một ảnh khuôn mặt vào không gian chính  $F$   
và phần bù trực giao  $\bar{F}$ .

Phương pháp thứ hai [263] dùng biệt số tuyến tính Fisher (Fisher's Linear Discriminant – FLD) để chiếu các mẫu từ không gian ảnh có số chiều cao sang một không gian đặc trưng có số chiều thấp hơn.



Và trên cơ sở phân tích biệt số tuyến tính, các tác giả đã xây dựng phương pháp Fisherface [148] và những phương pháp khác [249, 269] giải quyết tốt hơn phương pháp Eigenface [255] trong nhận dạng khuôn mặt. Khi dùng FLD để phân loại mẫu sẽ tốt hơn PCA khi chiếu. Do đó, kết quả phân loại trong không gian con đã được chiếu có thể khá hơn các phương pháp khác (ở [213] trình bày rõ về kích thước tập huấn luyện). Trong phương thức thứ hai, các tác giả đã phân rã các mẫu huấn luyện khuôn mặt và không phải khuôn mặt vào trong vài lớp con bằng ánh xạ tự tổ chức Kohonen (Kohonen's Self Organizing Map – SOM) [199]. Hình 9 cho thấy một đại diện của mỗi lớp khuôn mặt. Từ các mẫu được gán nhãn lại, tính các ma trận các giá trị rời rạc về tính chất mẫu ở trong lớp hay ở giữa lớp, bằng cách đó phát sinh phép chiếu tối ưu trên cơ sở FLD. Mỗi nhóm con, mô hình hóa mật độ như một phương thức Gauss với các tham số trong Gauss được ước lượng bằng phương pháp cực đại hóa khả năng [167]. Quét trên toàn bộ ảnh đưa vào bằng một cửa sổ rồi tính xác suất mức độ phụ thuộc lớp. Dùng luật quyết định dựa trên cực đại hóa khả năng để xác định có phải là khuôn mặt hay không. Cả hai phương pháp trong [263] có tỷ lệ chính xác là 92.3% cho MFA và 93.6% khi dùng FLD.

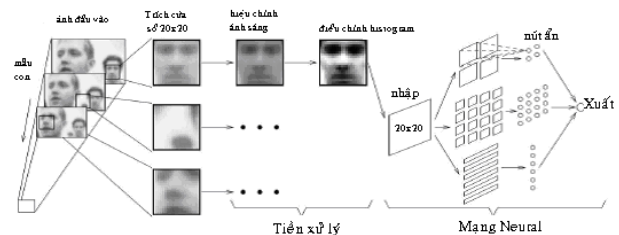


Hình 9: Đại diện của mỗi lớp khuôn mặt.  
Mỗi đại diện tương ứng tâm của một nhóm.

Choi [31] xây dựng hệ thống xác định khuôn mặt người trong ảnh màu bằng đặc trưng của mắt người thông qua phân đoạn để xác định ứng viên khuôn mặt dựa trên phân bố màu da của khuôn mặt.

### c) Mạng Neural

Mạng neural được áp dụng khá thành công trong các bài toán nhận dạng mẫu, như: nhận ký tự, đối tượng, robot tự vận hành. Xác định khuôn mặt người có thể xem là bài toán nhận dạng hai loại mẫu, có nhiều kiến trúc mạng neural đã được trình bày. Một thuận lợi khi dùng mạng neural để xác định khuôn mặt là tính khả thi của hệ thống học khi có sự phức tạp trong lớp của các mẫu khuôn mặt. Tuy nhiên, một điều trở ngại là các kiến trúc mạng đều tổng quát, khi áp dụng thì phải xác định rõ ràng số lượng tầng, số lượng node, tỷ lệ học, ..., cho từng trường hợp cụ thể, hình 10.



Hình 10: Mô hình mạng Neural theo Rowley

Agui trình bày mạng neural xử lý có thứ tự [143]. Giai đoạn đầu dùng hai mạng con song song mà dữ liệu vào là các giá trị cường độ của ảnh ban đầu và các giá trị cường độ của ảnh đã được lọc bằng thuật toán lọc Sobel với cửa sổ lọc 3x3. Đầu vào của mạng ở giai đoạn hai bao gồm dữ liệu đầu ra từ hai mạng con và các giá trị đặc trưng đã được trích ra, như: đặc trưng độ lệch chuẩn của các giá trị điểm ảnh trong mẫu đưa vào, một tỷ lệ của số điểm ảnh trắng trên tổng số điểm ảnh (đã chuyển sang ảnh nhị phân) trong một cửa sổ, và đặc trưng thiết yếu về hình học. Một giá trị xuất tại giai đoạn hai cho biết có tồn tại hay không khuôn mặt người trong vùng đưa vào. Qua kinh nghiệm, tác giả chỉ ra rằng nếu các ảnh cùng một kích thước thì mới dùng phương pháp này được.

Propp và Samal phát triển mạng neural để xác định khuôn mặt người sớm nhất [224]. Mạng neural của hai ông gồm bốn tầng với 1,024 đầu vào, 256 đầu kế tiếp trong tầng ẩn thứ nhất, tám đầu kế tiếp trong tầng ẩn thứ hai, và hai đầu ra. Tương tự như mạng neural xử lý theo thứ tự được đưa ra sau đó [251]. Phương pháp của Soulie [242] duyệt một ảnh đưa vào với mạng neural có thời gian trễ [258] (kích thước cửa sổ là 20x25 điểm ảnh) để xác định khuôn mặt. Dùng biến đổi wavelet để phân rã ảnh các phần

có kích thước khác nhau để xác định khuôn mặt. Vaillant dùng mạng neural dạng xoắn để xác định khuôn mặt người [256]. Đầu tiên tạo các ảnh mẫu khuôn mặt và không phải khuôn mặt có kích thước 20x20. Dùng một mạng neural, mạng này đã được huấn luyện, để tìm các vị trí tương đối của các khuôn mặt ở các tỷ lệ khác nhau. Rồi dùng một mạng khác để xác định vị trí chính xác của các khuôn mặt. Mạng đầu tiên dùng để tìm các ứng viên khuôn mặt, rồi dùng mạng thứ hai để xác định ứng viên nào thật sự là khuôn mặt. Burel và Carel dùng mạng neural đa tầng có ít mẫu hơn với thuật toán Kohonen's SOM để học các mẫu khuôn mặt và hình nền, mà các mẫu này đã được phân loại trước. Giai đoạn xác định khuôn mặt bao gồm duyệt trên mỗi ảnh đã được biến đổi từ ảnh bản đầu ở các độ phân giải khác nhau. Ở tại mỗi vị trí và kích thước của sổ duyệt, điều chỉnh độ sáng. Mỗi cửa sổ đã được chuẩn hóa sẽ được phân loại bằng MLP.

Feraud và Bernier dùng mạng neural kết hợp tự động [171, 172, 173]. Ý tưởng dựa trên [201] mạng kết hợp tự động có năm tầng thì có thể biểu diễn một phân tích thành phần chính phi tuyến. Dùng một mạng kết hợp tự động để xác định các khuôn mặt chụp thẳng rồi mở rộng bằng cách xoay 60 độ từ trái sang phải khuôn mặt chụp thẳng, mạng này sẽ tận dụng các trọng số khi xây dựng với dữ liệu khuôn mặt chụp thẳng cho các tư thế mới. Hai ông cho biết kết quả cũng tương tự [231]. Phương pháp này cũng dùng cho LISTEN [159] và MULTRAK [149].

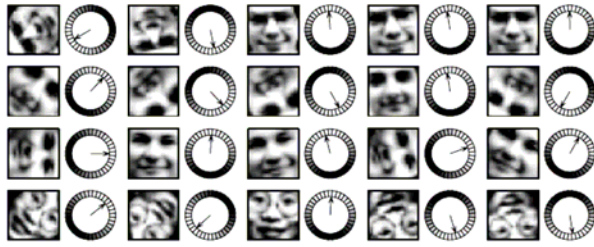
Lin xây dựng mạng neural quyết định trên cơ sở xác suất (Probabilistic Decision-based Neural Network – PDBNN) [209]. Kiến trúc của PDBNN thì tương tự một mạng có hàm trên nền tảng tương tự tia (Radial Basis Function – RBF) với các luật học được hỗ trợ xác suất. Thay vì chuyển toàn bộ ảnh khuôn mặt thành một vector có các giá trị cường độ để huấn luyện cho mạng neural, ông sẽ trích vector đặc trưng dựa trên cường độ và thông tin cạnh trong vùng khuôn mặt có chứa lông mày, mắt, và mũi. Hai vector đặc trưng đã được trích thì đưa vào hai PDBNN và hợp nhất các kết quả để có kết quả phân loại. Trên cơ sở 23 ảnh của Sung và Poggio [248]. Ông đã cho một số kết quả so sánh với các mạng khác [48, 248].

Theo đánh giá các phương pháp dùng mạng neural để xác định khuôn mặt người của nhiều tác giả, thì nghiên cứu của Rowley [48, 231] được xem là tốt nhất đối với ảnh xám. Một mạng đa tầng được dùng để học các mẫu khuôn mặt và không phải khuôn từ các ảnh tương ứng (dựa trên quan hệ cường độ, về mặt không gian của các điểm ảnh) trong khi Sung [246] dùng mạng neural để xác định một hàm biệt số cho mục đích phân loại mẫu có phải là khuôn mặt hay không dựa vào độ đo khoảng cách. Hai ông cùng dùng nhiều mạng neural và vài phương thức quyết định để cải thiện kết quả, trong khi Burel và Carel [153] dùng một mạng đơn, và Vaillant [256] dùng hai mạng để phân loại. Có hai thành phần chính để xử lý: nhiều mạng neural (xác định mẫu nào là khuôn mặt) và một mô đun để quyết định (đưa ra quyết định cuối cùng từ nhiều kết quả xác định). Hình 9, thành phần đầu tiên của phương pháp này là một mạng neural nhận một vùng ảnh có kích thước 20x20 điểm ảnh và xuất ra một giá trị trong khoảng từ -1 đến 1. Khi đưa vào một ảnh, nếu kết quả gần -1 thì nghĩa là mẫu này không phải là khuôn mặt người, nhưng nếu kết quả gần 1 thì đây chính là khuôn mặt người. Để xác định khuôn mặt có kích thước lớn hơn 20x20 điểm ảnh, cứ chọn một tỷ lệ rồi duyệt rồi xác định, rồi lại thay đổi tỷ lệ (biến thiên tỷ lệ này do người xây dựng quyết định). Gần 1050 mẫu khuôn mặt có kích thước, hướng, vị trí, và cường độ khác nhau dùng để huấn luyện mạng. Sẽ gán nhãn cho mắt, đỉnh của mũi, góc cạnh, và tâm của miệng rồi dùng để chuẩn hóa khuôn mặt về cùng một tỷ lệ, hướng, và vị trí. Thành phần thứ hai là phương pháp trộn các xác định chồng chéo nhau và đưa ra quyết định. Phép toán logic (AND/OR) là một quyết định đơn giản nhất và phương pháp bầu cử được dùng để tăng tính hiệu quả. Rowley [48] đưa nhiều cách giải quyết bài toán quyết khác nhau nhưng chi phí tính toán ít hơn Sung và Poggio nhưng tỷ lệ chính xác cao hơn.

Một giới hạn của phương pháp của Rowley [48] và Sung [246] là chỉ có thể xác định khuôn mặt chụp thẳng và tựa thẳng (nghiêng đầu). Sau đó Rowley [49] cải tiến để có thể xác định khuôn mặt bị xoay bằng mạng định hướng (Router Network), hình 11, sẽ thêm tiến trình xác định hướng khuôn mặt và



xoay về lại tư thế chuẩn (chụp thẳng), tuy nhiên khi quay lại dữ liệu như trên thì tỷ lệ chính xác lại giảm đi, chỉ còn khoảng 76.9%.



Hình 11: Một ví dụ cho dữ liệu vào và dữ liệu ra của mạng định hướng.

Lee [71] phát triển ý tưởng của Rowley [48] cho xác định khuôn mặt trong ảnh màu. Ông dùng mô hình màu da người bằng Gauss để xác định các ứng viên, sau đó loại bớt những ứng viên nào không thỏa mãn tính chất hình dáng gần giống hình ellipse. Cuối cùng ông dùng một mạng neural đã được huấn luyện để xác định khuôn mặt người. Tỷ lệ xác định chính xác là 88.9%, còn tỷ lệ xác định sai là 11.1%.

Dựa trên nghiên cứu của Rowley [48], Hazem [108] cải tiến để tốc độ xử lý tăng lên đáng kể.

Kwalek [131] dùng bộ lọc Gabor để trích đặc trưng, dùng đặc trưng này để huấn luyện cho mạng neural xoắn. Mạng neural xoắn là mạng neural mà mỗi node ở mỗi tầng có thể liên kết với các láng giềng cục bộ tầng phía trước của nó. Tỷ lệ chính xác là 87.5%.

#### d) SVM

Support Vector Machine (SVM) đã được Osuna [220] áp dụng đầu tiên để xác định khuôn mặt người. SVM được xem như là một kiểu mới dùng huấn luyện để phân loại theo hàm đa thức. Trong khi hầu hết các phương pháp khác huấn luyện để phân loại (Mạng Bayes, Neural, RBF) đều dùng tiêu chí tối thiểu lỗi huấn luyện (rủi ro do kinh nghiệm), trong khi SVM dùng quy nạp (được gọi là tối thiểu rủi ro cấu trúc), mục tiêu là làm tối thiểu một bao bên trên trên lỗi tổng quát. Một phân loại SVM là một phân loại tuyến tính, dùng một siêu phẳng để tách dữ liệu. Dựa trên một kết hợp có các trọng số của một tập con nhỏ các vector huấn luyện, các vector này được gọi là support vector. Ước lượng siêu phẳng thì tương đương giải một bài toán tuyến tính bậc hai. Osuna [220] đã phát triển một phương pháp hiệu quả để huấn luyện một SVM với tỷ lệ lớn để áp dụng cho bài toán xác định khuôn mặt người.

Ông dùng 10,000,000 mẫu có kích thước 19x19 điểm ảnh, hệ thống của ông có tỷ lệ lỗi ít hơn Sung và Poggio [247], nhưng nhanh hơn gần 30 lần. SVM cũng có thể dùng xác định khuôn mặt người và người đi bộ với phân tích Wavelet [219, 221, 222].

Shihong và Masato sử dụng biến đổi wavelet Gabor để trích đặc trưng của khuôn mặt cũng như không phải khuôn mặt để đưa vào cho SVM học [15].

Kang và Lee [14] xây dựng ứng dụng cho robot đi bộ vượt qua con người và chướng ngại vật dựa trên xác định khuôn mặt. Hai ông dùng phân đoạn nổi kết hợp SVM để phân loại. Tương tự Kui và Silva [22] cũng xây dựng ứng dụng cho phòng thông minh bằng cách xác định khuôn mặt người dựa trên eigenface để làm dữ liệu cho SVM học phân loại.

Bileschi và Heisele [18] dùng độ phân giải thấp để học thành phần khuôn mặt trong ảnh xám với các khuôn mặt chụp thẳng hoặc tựa thẳng cho SVM để xác định khuôn mặt. Trong khi Terrillon [20] dùng tính chất màu da người để tìm ứng viên kết hợp SVM và các mô men Fourier-Mellin trực giao để giải quyết. Thay vì lọc đơn giản, Shu-Fai và Kwan-Yee [72] dùng QuaTree để tìm ứng viên khuôn mặt người trong ảnh màu. Sau đó kết hợp wavelet phân tích mẫu cho SVM học trong nhiều tỷ lệ. Đa phần khi cho SVM học, các tác giả đều dùng hai lớp khuôn mặt và không phải khuôn mặt để học. Wang [75] chỉ dùng một lớp khuôn mặt trong ảnh màu để xác định khuôn mặt người. Tỷ lệ chính xác khoảng 81%. Fang và Qiu [83] kết hợp SVM và thuật toán leo đồi để xác định khuôn mặt. Zhang và Zhao [51] xây dựng SVM dựa trên histogram của khuôn mặt và không phải khuôn mặt để xác định khuôn mặt. Tỷ lệ chính xác khoảng 92% cho khuôn mặt chụp thẳng hoặc gần thẳng trong ảnh màu. Je lại xây dựng nhiều SVM để xác định khuôn mặt người theo thứ tự quyết định kết hợp phương pháp bầu cử trong ảnh màu [30].

Julien [129] xây dựng một cấu trúc SVM mới gồm nhiều SVM kết nối song song với nhau học dữ liệu từ không gian eigenface. Tỷ lệ chính xác hơn 93% trong ảnh xám với khuôn mặt đơn được chụp thẳng.

e) *Mạng lọc thưa*

Yang đề xuất một phương pháp dùng mạng lọc dư thừa (Sparse Network of Winnows – SNoW) [181, 230] để xác định khuôn mặt người với các đặc trưng khác nhau và biểu diễn trong các tư thế khác nhau, và dưới điều kiện ánh sáng khác nhau [264]. Đồng thời nghiên cứu phương pháp học sơ khai tốt như thế nào khi dùng các đặc trưng đa tỷ lệ. SNoW là một mạng thưa dùng các hàm tuyến tính và dùng lọc để cập nhật luật [210]. Phương pháp này thích hợp cho học trong miền khi các đặc trưng tiềm năng tạo các quyết định sai khác nhau mà không biết mức độ ưu tiên. Một vài đặc tính của kiến trúc học này là rất hiếm dữ liệu có phân chung, có chỉ định các đặc trưng và liên kết trong dữ liệu, kỹ thuật quyết định, và cập nhật luật hiệu quả. Tỷ lệ lỗi là 5.9%, hiệu quả cũng như các phương pháp khác [48, 160, 220, 237].

Gundimada [4] dựa trên kiến trúc SNoW để xây dựng ba mạng, mạng thứ nhất phân loại dựa trên phân bố cường độ, hai mạng dựa trên phân bố màu da người để tìm ứng viên kết hợp phương pháp làm nổi bật cạnh. Xây dựng các mẫu đầy đủ tư thế của khuôn mặt, mỗi bộ phân loại sẽ tương ứng cho một hướng, mỗi hướng lệch nhau  $10^\circ$ .

f) *Phân loại Bayes*

Trái ngược với các phương pháp trong [48, 220, 248] dựa vào diện mạo trên toàn khuôn mặt, Schneiderman và Kanade mô tả một phân loại naive Bayes để ước lượng xác suất để nổi các diện mạo tại vị trí cục bộ trên khuôn mặt và vị trí của các mẫu khuôn mặt người (các vùng con trên khuôn mặt) trong nhiều độ phân giải [73, 237]. Hai ông nhấn mạnh tính chất diện mạo khuôn mặt ở vị trí cục bộ bởi vì với vài mẫu ở vị trí cục bộ của một đối tượng sẽ có tính chất duy nhất, cường độ xung quanh mẫu mất thì đặc biệt hơn ở vị trí má. Đây là hai lý do để dùng phân loại naive Bayes (không xem xét thống kê những phụ thuộc giữa các vùng). Đầu tiên phân loại này cung cấp ước lượng tốt hơn của các hàm mật độ có điều kiện của các vùng này. Thứ hai, một phân loại Bayes cung cấp một dạng hàm của theo xác suất để nhận thống kê của diện mạo ở vị trí cục bộ và vị trí của nó trên đối tượng. Tại mỗi tỷ lệ, một ảnh khuôn mặt người được phân rã làm bốn vùng hình chữ nhật con. Chiếu các vùng này xuống không

gian có số chiều thấp hơn (dùng PCA để xây dựng) và lượng tử hóa thành một tập các mẫu có giới hạn, và thống kê mỗi vùng đã được chiếu, các thống kê này được ước lượng từ các mẫu được chiếu xuống không gian có số chiều nhỏ hơn, để mã hóa diện mạo cục bộ. Khi tỷ lệ khả năng lớn hơn tỷ lệ của các xác suất ưu tiên thì có khuôn mặt người. Ông cũng cho thấy so sánh giữa phương pháp này và [48], hướng tiếp cận này cho phép xác định các khuôn mặt bị xoay và nhìn nghiêng. Schneiderman và Kanade sau đó kết hợp biến đổi wavelet để xác định các khuôn mặt nhìn nghiêng và xe hơi [58].

Rickert cũng dùng cách chọn các đặc trưng cục bộ [229]. Các đặc trưng cục bộ được trích ra bằng cách áp dụng các bộ lọc đa tỷ lệ và nhiều độ phân giải trên dữ liệu ảnh đưa vào. Dùng phương pháp gom nhóm dữ liệu và một Gauss hỗn hợp để tìm phân bố của các vector đặc trưng. Sau khi huấn luyện cho mô hình và tính chế, tính khả năng của các vector đặc trưng của các ảnh để phân loại. Phương pháp này cho kết quả tốt cho xác định khuôn mặt người cũng như xe hơi.

Thang [77] xác định khuôn mặt người thông qua phân loại mạng Bayes kết hợp, hay còn gọi là mạng Bayes có cấu trúc như rừng (Forest-Structured Bayesian Network), để xác định các biệt số. Kết hợp phương pháp Bagging để xây dựng phân loại tích hợp nhằm xác định khuôn mặt người trong ảnh xám. Tỷ lệ chính xác hơn 90%.

Nam và Rhee [110, 123] xây dựng mạng Bayes học phân loại theo ngữ cảnh: màu da, ánh sáng, và kết cấu khuôn mặt và kết hợp mạng FuzzyART để xác định khuôn mặt người trong ảnh. Hai tác giả này cũng dùng thêm khoảng cách Mahalanobis [122] khi kết hợp mạng RBF và FuzzyART để xác định khuôn mặt có nhiều tỷ lệ khác nhau. Hai tác giả phát triển bằng cách dùng nhiều phân loại Bayes để chọn ứng viên thông qua các đặc trưng thông tin về cường độ và kết cấu của khuôn mặt [126]. Tỷ lệ chính xác hơn 87%.

Lee và Kim [120] dùng đặc trưng Haar wavelet 1-chiều để huấn luyện cho mạng Bayes để xác định nhiều khuôn mặt chụp thẳng trong ảnh xám thông qua PDF của các mẫu khuôn mặt người và mẫu

không phải khuôn mặt người. Tỷ lệ chính xác là 98%.

Zhu [97] dùng wavelet để trích các tham số đặc trưng dựa vào histogram rồi dùng mạng Bayes đã được học để xác định khuôn mặt người trong nhiều tỷ lệ khác nhau.

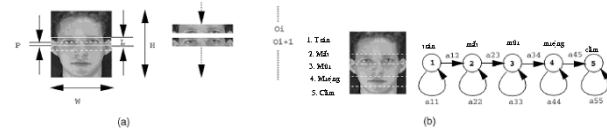
Duy Nguyen [280] dùng bộ lọc Sobel để xác định các đặc trưng rồi dùng phân loại naive Bayes như Schneiderman và Kanade để xác định khuôn mặt người.

#### g) Mô hình Markov ẩn

Một giả thuyết quan trọng của mô hình Markov ẩn (Hidden Markov Model – HMM) là các mẫu có thể được đặc tính hóa như các tiến trình ngẫu nhiên có tham số và các tham số này được ước lượng chính xác, đây là một trong những định nghĩa rõ ràng. Khi phát triển HMM để giải quyết bài toán nhận dạng mẫu, phải xác định rõ có bao nhiêu trạng thái ẩn đầu tiên cho hình thái mô hình. Sau đó, huấn luyện HMM học xác suất chuyển tiếp giữa các trạng thái từ các mẫu, mà mỗi mẫu được mô tả như một chuỗi các quan sát. Mục tiêu huấn luyện HMM là cực đại hóa xác suất của quan sát từ dữ liệu huấn luyện bằng cách điều chỉnh các tham số trong mô hình HMM thông qua phương pháp phân đoạn Viterbi chuẩn và các thuật toán Baum-Welch [227]. Sau khi huấn luyện xong, dựa vào xác suất để xác định một quan sát thuộc lớp nào.

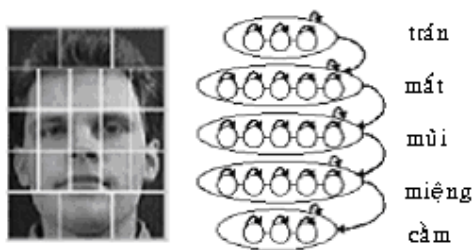
Một cách trực quan, có thể chia một mẫu khuôn mặt người thành nhiều vùng khác nhau như đầu, mắt, mũi, miệng, và cằm. Có thể nhận dạng một mẫu khuôn mặt người bằng một tiến trình xem xét các vùng quan sát theo một thứ tự thích hợp (từ trên xuống dưới, từ trái qua phải). Thay vì tin tưởng vào mức độ chính xác vị trí lẽ để dùng cho các phương pháp dựa trên so khớp hay dựa trên diện mạo (nơi xuất hiện các đặc trưng như mắt và mũi cần xác định vị trí lẽ tốt để lấy được toàn bộ chi tiết của đặc trưng). Mục tiêu của hướng tiếp cận này là kết hợp các vùng đặc trưng khuôn mặt với các trạng thái của mô hình. Thường các phương pháp dựa vào HMM sẽ xem xét một mẫu khuôn mặt như một chuỗi các vector quan sát, với mỗi vector là một dãy các điểm ảnh, hình 12a và hình 13. Trong quá trình huấn luyện và kiểm tra, một ảnh được quét theo một thứ

tự và một quan sát được xem như một khối các điểm ảnh, hình 12a và hình 13. Áp dụng một định hướng theo xác suất để chuyển từ trạng thái này sang trạng thái khác, hình 12b, dữ liệu ảnh được mô hình hóa bằng phân bố Gauss nhiều biến. Một chuỗi quan sát bao gồm tất cả giá trị cường độ từ mỗi khối. Kết quả xuất ra cho biết quan sát thuộc lớp nào. HMM được dùng để nhận dạng khuôn mặt người và xác định khuôn mặt người. Samaria [235] dùng năm trạng thái tương ứng năm vùng, hình 12b để mô hình hóa tiến trình xác định khuôn mặt người. Ông huấn luyện từng vùng cho HMM. Mỗi tình trạng sẽ phụ trách xem xét vùng tương ứng để đưa ra quyết định phù hợp. Nếu kết quả xem xét cuối cùng vượt qua một ngưỡng thì quan sát này sẽ là khuôn mặt người.



Hình 12: Mô hình Markov ẩn:  
(a) các vector quan sát để huấn luyện cho HMM;  
(b) năm trạng thái ẩn.

Samaria và Young dùng HMM 1-chiều (hình 12) và 2-chiều (hình 13) để trích đặc trưng khuôn mặt dùng để nhận dạng khuôn mặt [234, 235]. HMM khai thác cấu trúc của khuôn mặt tuân theo các chuyển tiếp trạng thái. Từ các vùng có đặc trưng quan trọng như: tóc, trán, mắt, mũi, và miệng, hai ông phân tích theo tự nhiên từ trên xuống dưới, mỗi vùng được thiết kế thành một trạng thái 1-chiều. Mỗi ảnh được phân đoạn chuẩn thành năm vùng theo thứ tự từ trên xuống dưới tạo thành năm trạng thái. Hai ông dùng phân đoạn Viterbi thay cho phân đoạn chuẩn và các tham số trong HMM được tái ước lượng bằng thuật toán Baum-Welch. Tương tự [234], Nefian và Hayes dùng HMM và biến đổi Karhunen Løeve (Karhunen Løeve Transform – KLT) để xác định khuôn mặt người và nhận dạng [217]. Thay vì dùng các giá trị cường độ thô, các vector quan sát sẽ bao gồm các hệ số (dùng KLT để có) thì kết quả sẽ tốt hơn [234], và tỷ lệ chính xác khi dùng HMM 2-chiều (hình 13) là 90%.



Hình 13: Xác định khuôn mặt bằng HMM các trạng thái, mỗi trạng thái lại có những trạng thái nhỏ bên trong: trạng thái trán có ba trạng thái nhỏ bên trong; trạng thái mắt có năm trạng thái nhỏ bên trong.

Rajagopalan đưa ra hai phương pháp xác suất để xác định khuôn mặt người [228]. Tương phản với [248], dùng một tập các Gauss nhiều biến để mô hình hóa phân bố của khuôn mặt người, phương pháp đầu tiên trong [228] dùng tổng kê có thứ tự ở mức cao hơn (Higher Order Statistic - HOS) để ước lượng cường độ. Tương tự [248], các phân bố không biết của khuôn mặt và không phải khuôn mặt được gom nhóm bằng sáu hàm cường độ dựa trên HOS của các mẫu. Như trong [246], sử dụng tri giác nhiều tầng để phân loại, một vector đưa vào xử lý gồm mười hai đo lường khoảng cách giữa mẫu ảnh và mười hai nhóm. Tiếp cận này dựa trên cơ sở sinh một dãy quan sát từ ảnh rồi dùng HMM học các tham số tương ứng. Kết quả của ông cho thấy cả hai phương pháp HOS và HMM đều có kết quả xác định khuôn mặt người cao hơn [48, 248], nhưng nhiều xác định nhầm hơn.

Filareti dùng đặc trưng sắc màu kết hợp thông tin về độ sâu của ảnh làm dữ liệu đầu vào dạy cho HMM để xác định khuôn mặt người [63]. Phương pháp này cho phép giải quyết vấn đề về điều kiện hình nền, độ sáng, che khuất, tư thế khuôn mặt.

Hong [121] xây dựng mô hình Markov ẩn học dữ liệu dựa trên các đặc trưng Haar-like để xác định khuôn mặt người. Tỷ lệ chính xác là 96%.

#### *h) Hướng tiếp cận lý thuyết thông tin*

Thuộc tính trong không gian của mẫu khuôn mặt có thể được mô hình hóa qua nhiều diện mạo khác nhau. Dùng ngữ cảnh để phân đoạn là một phương pháp hiệu quả, xác định ngữ cảnh thông qua các điểm ảnh lân cận. Lý thuyết trường ngẫu nhiên Markov (Markov Random Field – MRF) cung cấp một tiện lợi và cách phù hợp để mô hình hóa các thực thể dựa vào ngữ cảnh như các điểm ảnh và các đặc trưng có mối tương quan. Theo định lý

Hammersley-Clifford, một MRF có thể được đặc tính hóa tương đương bằng một phân bố Gibbs và các tham số thường cực đại hóa sau khi ước lượng [225]. Như một sự lựa chọn, các phân khuôn mặt người và không phải khuôn mặt có thể được ước lượng qua các histogram. Dùng thông tin quan hệ Kullback, tiến trình Markov cực đại hóa biệt số trên cơ sở thông tin giữa các lớp để xác định khuôn mặt người [160, 207].

Lew áp dụng thông tin quan hệ Kullback [162] kết hợp hàm xác suất  $p(x)$  khi mẫu là khuôn mặt người và  $q(x)$  khi mẫu không phải là khuôn mặt người để xác định khuôn mặt người [207]. Ông dùng 100 cá thể khuôn mặt người gồm chín quang cảnh để ước lượng phân bố của khuôn mặt. Dùng 143,000 mẫu không phải khuôn mặt để ước lượng hàm mật độ xác suất (PDF) thông qua histogram. Từ đây chọn được các điểm ảnh giàu thông tin nhất (Most Informative Pixel – MIP) để cực đại hóa thông tin quan hệ Kullback giữa  $p(x)$  và  $q(x)$  (có được một phân tách lớp cực đại). Phân bố MIP tập trung ở các vùng mắt và miệng, nhưng bỏ qua vùng mũi. MIP được dùng để có được các đặc trưng tuyến tính dùng cho phân loại và mô tả bằng phương pháp của Fukunage và Koontz [175]. Dùng một cửa sổ duyệt trên toàn bộ ảnh để xây dựng khoảng cách từ không gian khuôn mặt (Distance From Face Space – DFFS), được định nghĩa ở [283]. Nếu DFFS đến không gian con khuôn mặt ngắn hơn khoảng cách đến không gian con không phải khuôn mặt, hình 8, thì xem như đã xác định được khuôn mặt trong cửa sổ.

Thông tin quan hệ Kullback cũng được Colmenarez và Huang dùng để cực đại hóa biệt số trên cơ sở thông tin giữa các mẫu negative và positive của khuôn mặt [160]. Phân tích các ảnh từ tập huấn luyện của mỗi lớp (lớp khuôn mặt người và lớp không phải khuôn mặt người) như các quan sát trong tiến trình ngẫu nhiên và được đặc tính hóa bằng hai hàm xác suất. Hai ông dùng một học các quá trình xử lý Markov rời rạc để mô hình các mẫu khuôn mặt và hình nền rồi ước lượng mô hình xác suất tương ứng. Quá trình học được chuyển thành bài toán tối ưu để chọn được tiến trình cực đại biệt số trên cơ sở thông tin giữa hai lớp. Tính tỷ lệ khả

năng dùng cho mô hình xác suất đã được huấn luyện rồi dùng để xác định khuôn mặt người.

Qian và Hang [225] trình bày một phương pháp dùng cả hai phương pháp trên cơ sở quang cảnh và mô hình hóa. Đầu tiên, một thuật toán dùng tri thức miền ở mức cao của những gì khi nhìn vào thì quan tâm ngay để giảm số chiều không gian tìm kiếm (thay vì tìm trên toàn bộ không gian đặc trưng, chỉ cần tìm trên không gian con có những đặc trưng quan tâm). Thuật toán này chọn các vùng trên ảnh làm mục tiêu khi có diện mạo quan tâm xuất hiện bằng thuật toán xác định vùng (phương pháp watershed). Dùng các vùng đã được chọn để xác định khuôn mặt so khớp mẫu và so khớp đặc trưng thông qua trường ngẫu nhiên Markov và cực đại ước lượng sau.

Feng và Shi [102] dùng KFD (Kernel Fisher Discriminant) để phân tích ảnh có khuôn mặt người rồi học các đặc trưng này để xác định khuôn mặt người. Tỷ lệ chính xác trong ảnh xám khoảng 72%.

#### i) Học theo quy nạp

Các thuật toán học theo quy nạp được áp dụng để xác định khuôn mặt người. Huang dùng thuật toán C4.5 [226] xây dựng cây quyết định từ các mẫu khuôn mặt người [187]. Mỗi mẫu huấn luyện là một cửa sổ có kích thước 8x8 điểm ảnh và được mô tả như một vector có 30 thuộc tính về entropy, trung bình, và độ lệch chuẩn của các giá trị cường độ của điểm ảnh. Mỗi node của cây quyết định sẽ chỉ rõ quyết định trên một thuộc tính đơn. Tỷ lệ chính xác khi xác định khuôn mặt được chụp thẳng là 96%.

Duta và Jain [273] mô tả một phương pháp học khái niệm khuôn mặt người bằng thuật toán Find-S của Mitchell [215]. Tương tự [248], hai ông ước đoán phân bố của các mẫu khuôn mặt người bằng  $p(x | \text{khuôn mặt})$  thông qua một tập các nhóm theo Gauss và khoảng cách từ mẫu khuôn mặt đến tâm nhóm nên nhỏ hơn khoảng cách lớn nhất từ các điểm đến tâm nhóm (nghĩa là name trong phần bao của nhóm). Sau đó dùng thuật toán Find-S học khoảng cách ngưỡng. Phương pháp này có vài đặc tính riêng. Thứ nhất, không dùng các mẫu không phải là khuôn mặt người, trong khi [48, 248] dùng cả hai loại mẫu. Thứ hai, chỉ dùng duy nhất phần tâm để huấn luyện. Thứ ba, các vector đặc trưng gồm có các

ảnh với 32 mức cường độ hoặc kết cấu, trong khi [248] dùng toàn bộ tỷ lệ các giá trị cường độ. Tỷ lệ chính xác là 90%.

Bernhard Froba và Andreas Ernts [25] dùng cây quyết định có nhiều nhánh cho phép xác định khuôn mặt người nhìn nghiêng từ  $-60^\circ$  đến  $60^\circ$ , mỗi node có khả năng loại bỏ cửa sổ con hiện hành đang xét hoặc phân loại vào một trong ba lớp quay. Tỷ lệ chính xác cho ảnh xám là 90%.

Socolinsky [91] dùng phân loại CCCD (Class-Cover Catch Digraph) kết hợp boosted tree-like thông qua độ đo cross-correlation để xác định khuôn mặt người dựa trên tập mẫu huấn luyện.

Ramana [112] dùng cây quyết định như một công cụ để phân loại xem phần nào sẽ là khuôn mặt người. Trong khi xây dựng cây ông kết hợp cả cascade để tăng tính hiệu quả.

#### j) AdaBoost

Học với AdaBoost là một phân loại mạnh phi tuyến phức  $H_M(x)$ , được xây dựng từ  $M$  phân loại

yếu [274],  $H_M(x) = \frac{\sum_{m=1}^M \alpha_m h_m(x)}{\sum_{m=1}^M \alpha_m}$  với  $x$  là mẫu cần

phân loại,  $h_m(x) \in \{-1, 1\}$  là  $M$  phân loại yếu,  $\alpha_m \geq 0$  là các hệ số trong  $\square$ , và  $\sum_{m=1}^M \alpha_m$  là nhân tố để chuẩn hóa. Mục tiêu của Adaboost là học một dãy các phân loại yếu. Giả sử có một tập  $N$  mẫu huấn luyện đã được gán nhãn  $\{(x_1, y_1), \dots, (x_N, y_N)\}$ , với  $y_i$  là nhãn tương ứng của mẫu  $x_i \in \square^n$ . Tính một phân bố của các mẫu huấn luyện  $[w_1, \dots, w_N]$  cập nhật trong suốt quá trình học. Sau bước lặp  $m$ , mẫu khó phân loại  $(x_i, y_i)$  có trọng mới  $w_i^{(m)}$ , đến bước lặp thứ  $(m+1)$ , mẫu này sẽ có tầm quan trọng hơn.

Viola và Jones dùng AdaBoost kết hợp cascade để xác định khuôn mặt người [52] với các đặc trưng dạng Haar wavelet-like [221, 275]. Tốc độ xử lý khá nhanh và tỷ lệ chính xác hơn 80% trên ảnh xám.

Schneiderman và Kanade dùng wavelet để trích đặc trưng rồi xây dựng hệ thống học với Adaboost, dựa trên xác suất và histogram để xác định khuôn mặt người [26]. Tỷ lệ chính xác trên 90%.

Chen [37] ước lượng tham số ảnh để điều chỉnh ánh sáng cho phù hợp cho các mẫu bằng SVM. Sau

cùng dùng Adaboost để xác định khuôn mặt người với tư thế chụp thẳng. Tỷ lệ chính xác là 89.7%.

Lu [58] dùng phân tích biệt số tuyến tính (Linear Discriminant Analysis – LDA) để trích đặc trưng. Sau đó huấn luyện bộ phân loại yếu boosting để phân loại khuôn mặt người.

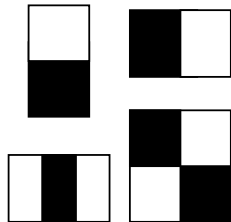
Shinji và Osamu [137] xây dựng các trung của khuôn mặt bằng cách sử dụng nhiều mức độ phân giải thấp để xác định khuôn mặt người thông qua Adaboost.

Jin [113] chỉ ra nếu dùng từng phương pháp so khớp mẫu hay cascade riêng rẽ thì mức độ chính xác gần như nhau, nhưng mức độ xác định sai khá cao. Tác giả kết hợp hai phương pháp này để giảm tỷ lệ sai của phương pháp xác định khuôn mặt người.

Ou [115] thấy rằng khi dùng cascade AdaBoost để xác định khuôn mặt người thì thông thường dùng thuật toán greedy để tìm các trọng của bộ phân loại yếu thì không được tối ưu. Tác giả đề xuất dùng GA để thay thế cách tìm trên nhằm tăng tính hiệu quả.

#### k) Các đặc trưng Haar-like và phân loại với cascade

Viola và Jones dùng bốn loại đặc trưng Haar-like cơ bản để xác định khuôn mặt người [52, 221], hình 13. Đặc trưng Haar được ưa thích vì có hai lý do: (1) phân loại mạnh trong việc xác định khuôn mặt người hay không phải khuôn mặt người; và (2) có hiệu quả [276] khi dùng bảng tổng các vùng [284] hoặc kỹ thuật ảnh đầy đủ [52].



Hình 13: Bốn loại đặc trưng Haar wavelet-like.

Ảnh đầy đủ  $I(x,y)$  tại vị trí  $(x,y)$  là tổng các điểm ảnh ở trên và bên trái của  $(x,y)$  [52],

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

Lienhart phát triển các đặc trưng Haar-like thành một bộ đặc trưng mới [19] kết hợp phân loại cascade để xác định khuôn mặt người.

Lin [79] dùng boost để loại trường hợp học quá khớp đồng thời sử dụng huấn luyện tăng cường để

giảm tỷ lệ sai. Cuối cùng dùng cascade để xử lý khuôn.

Maeynet và Thiran [62] xác định đặc trưng Gauss không đẳng hướng và Haar-like kết hợp Adaboost để xác định khuôn mặt người.

Tương tự Song [99] cũng dùng AdaBoost để xác định khuôn mặt người kết hợp màu da người. Rồi dùng các thông tin này để theo vết các đối tượng là con người ở trong phòng.

Ishii [64] xây dựng phương pháp BDF (Block Difference Feature) với các đặc trưng Haar-like để xác định các tư thế của khuôn mặt, rồi dùng kết quả này để so sánh với dữ liệu đã được học trước đó và xác định có khuôn mặt hay không trong ảnh.

#### l) Học với FloatBoost

Li và Zhang đưa ra một khái niệm mới đó là FloatBoost [103]. Phương pháp này học dựa trên phân loại boosting để tỷ lệ lỗi cực tiểu. Nhưng phương pháp này cho phép quay lui sau khi tại mỗi bước khi học bằng AdaBoost đã cực tiểu được tỷ lệ lỗi trực tiếp, cực tiểu theo hàm mũ. Có hai vấn đề gặp khi dùng phương pháp AdaBoost:

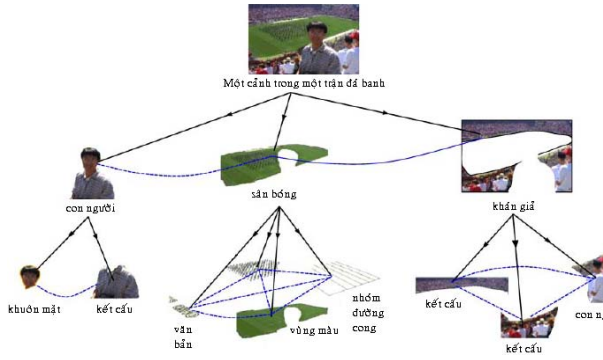
- *Thứ nhất:* AdaBoost cực tiểu theo hàm mũ tại biên qua tập huấn luyện. Đây là tiện lợi, tuy nhiên mục tiêu cuối cùng trong các ứng dụng dùng phân loại mẫu thì thường là cực tiểu một giá trị trực tiếp (tuyến tính) kết hợp với tỷ lệ lỗi. Một phân loại mạnh được học bằng AdaBoost thì gần điểm tối ưu của ứng dụng trong điều kiện tỷ lệ lỗi. Vấn đề này không thấy tài liệu nói đến có lời giải.
- *Thứ hai:* AdaBoost để lại một thách thức nếu dùng phân loại yếu để học. Học để phân loại tối ưu khi dùng phân loại yếu cần ước lượng mật độ không gian đặc trưng, điều này là vấn đề khó, đặc biệt khi số chiều của không gian khá lớn.

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

Tian [101] xây dựng cây xác định trên cơ sở học tích cực bằng thuật toán gom nhóm c-mean mở dựa



trên nền tảng FloatBoost. Tác giả dùng các đặc trưng Harr wavelet-like để áp dụng cho phương pháp của mình. Phương pháp này cho phép xác định nhiều khuôn mặt người ở nhiều tư thế khác nhau trong ảnh màu.



Hình 15: Một ví dụ về cú pháp ảnh.

#### m) Phân loại dựa trên cú pháp hay mảnh

Tu [141] dựa trên khái niệm cú pháp trong xử lý ngôn ngữ để xây dựng đồ thị cú pháp của ảnh dựa trên nội dung ảnh, hình 15, với chuỗi Markov. Sau khi có được các từ vựng, ông dùng phương pháp Adaboost đã được huấn luyện trước đó để xác định các đối tượng.

Hoặc có thể xem từng vùng của khuôn mặt người như các phần cấu thành một khuôn mặt, rồi xây dựng các giả thuyết để xác định khuôn mặt người. Bastian [278] đã xây dựng một lý thuyết tổng quát cho các loại đối tượng.

#### n) Phân loại dựa trên loại bỏ

Elad xây dựng một phân loại dựa trên khái niệm loại bỏ tối đa (Maximal Rejection Classifier – MRC) khác hẳn ý tưởng phân loại khác. Các phương pháp khác tìm mức độ chung của một các thể nào đó so với các lớp để chọn cá thể đó vào lớp nào. Ông chọn cách loại bỏ những lớp mà cá thể này không có hoặc có ít mối tương quan, chi phí loại bỏ không cao lắm [28]. Ông tính PDF của hai lớp: khuôn mặt người và không phải khuôn mặt người. Ông xem khuôn mặt người là target và không phải khuôn mặt người là clutter. Ông tìm ngưỡng loại bỏ theo xác suất thông qua biệt số tuyến tính Fisher (FLD). Ông chiếu dữ liệu xuống một vector chiếu, thông qua phép chiếu này để xác định khuôn mặt người.

#### o) Hướng tiếp cận tổng hợp

Các các phương pháp được chia làm bốn phân loại chính theo bốn hướng tiếp cận. Tuy nhiên, có nhiều phương pháp không hoàn toàn rơi vào một trong bốn hướng tiếp cận này mà ở trong nhiều hướng tiếp cận khác nhau. Ví dụ, phương pháp so khớp mẫu dùng mô hình khuôn mặt người và các mẫu con để trích các đặc trưng khuôn mặt [163, 177, 232, 238, 269], và sau đó dùng các đặc trưng này để xác định khuôn mặt. Hơn nữa phương pháp dựa trên tri thức và phương pháp so khớp mẫu không thật sự tách biệt, từ đó có nhiều hướng giải quyết dùng tri thức của con người để định nghĩa các mẫu khuôn mặt người [164, 232, 238].

Kim [24] kết hợp các đặc trưng láng giềng của khuôn mặt để xây dựng các mẫu theo các hướng, sau đó dùng kỹ thuật xác định cạnh EBM (Edge-like Blob Map) theo cường độ. Ông xây dựng logic mờ kết hợp PCA để ước lượng tư thế các khuôn mặt.

Taur và Tao [68] xây dựng phân loại neuro-fuzzy (neuro-fuzzy classifier – NEFCAR) có độ đo tin cậy để biết ảnh nào là khuôn mặt người. Các ứng viên được chọn thông qua phân đoạn màu da.

Chen sử dụng MRF kết hợp hình thái học để xác định khuôn mặt người [87]. Ông dùng các bộ lọc top-hat, bottom-hat, và watershed để phân đoạn ảnh đối tượng đang di chuyển nhằm tìm ứng viên khuôn mặt. Sau cùng dùng MRF và hình dạng ellipse của khuôn mặt để xác định ứng viên nào là khuôn mặt.

Garcia và Tziritis [2, 11] sau khi lọc những vùng nào có màu là màu da người để tìm ứng viên, sau đó dùng thuật toán trộn vùng để tạo ứng viên mịn hơn. Hai ông dùng phân tích theo wavelet để phân rã ứng viên xem có cùng kết cấu với khuôn mặt người hay không thông qua khoảng cách Bhattacharyya. Tỷ lệ chính xác khoảng 95%.

Cooray và Connor [3] và Lee [36] dùng kỹ thuật lai trên cơ sở trích đặc trưng và PCA để xác định khuôn mặt người. Hai ông dùng các đặc trưng về màu sắc của mắt và miệng để xác định các vùng này. Từ nay áp dụng lý thuyết eigenface để chuẩn hóa không gian tìm kiếm, đồng thời dùng khoảng cách giữa hai mắt để có kết quả cuối cùng.

Zhang [41] kết hợp mô hình màu da người để phân đoạn tìm ứng viên khuôn mặt. Ông xây dựng mạng neural như của Rowley [49] để quay khuôn

mặt sau đó so khớp các mẫu có sẵn. Phương pháp này cho xác định các khuôn mặt ở các tư thế khác nhau trong ảnh màu, thời gian xử lý sẽ giảm hơn vì không gian tìm kiếm đã bị thu hẹp. Tương tự Haizhou [40] cũng dùng phương pháp như thế nhưng thay đổi quá trình xác định. Ông so khớp mẫu dùng để tìm ứng viên. Sau đó dùng mạng neural để phân loại ứng viên nào là khuôn mặt người.

Li [138] dùng kernel để học như là một ánh xạ phi tuyến, đầu tiên ông dùng KPCA (Kernel PCA) để chọn các đặc trưng và không gian đặc trưng để học. Sau đó ông dùng KSVC (Kernel Support Vector Classifier) kết hợp FLD để phân loại đầu là khuôn mặt người.

Yin và Meng [107] dùng phân đoạn ảnh tính chất màu da người, từ đây tác giả tìm được vị trí của mắt theo tiêu chuẩn mức độ cân đối để xem các vùng này là các ứng ứng của khuôn mặt. Từ các ứng viên này, tác giả so khớp với các mẫu có sẵn để xác định ứng viên nào là khuôn mặt người. Tỷ lệ chính xác là 85%.

Lingmin [93] và Emanuele [142] dùng lý thuyết học theo xác suất để xây dựng mô hình xác định khuôn mặt người. Trong khi Emanuele dùng đặc trưng Haar wavelet để xác định đặc trưng rồi dạy cho SVM để phân loại. Lingmin dùng ba lớp: tầng cuối dùng mối tương quan hình học các mẫu, tầng giữa xét tính đa dạng của các thành phần khuôn mặt người với HMM, và tầng trên cùng dùng mô hình đồ thị cho quan hệ ba thành phần tạo thành một hình tam giác (xét tính cân đối).

### III. KHÓ KHĂN VÀ THÁCH THỨC TRONG BÀI TOÁN XÁC ĐỊNH KHUÔN MẶT NGƯỜI

Việc xác định khuôn mặt người có những khó khăn nhất định [57] như sau, hình 16:

- 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 toạ độ của máy ảnh so với ảnh.
- Kích thước khác nhau của các khuôn mặt người, và đặc biệt là trong cùng một ảnh.
- Màu sắc của môi trường xung quanh, hay màu sắc quần áo của người được chụp lấy ảnh.
- Xuất hiện thành phần khuôn mặt hay không.
- 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 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*).

### IV. KẾT LUẬN

Bài viết này cố gắng cung cấp một cái nhìn tổng quan các phương pháp xác định khuôn mặt người và một cách phân loại các hướng tiếp cận từ ảnh xám đến ảnh màu dựa vào gần 300 bài báo, báo cáo của các phòng nghiên cứu, và luận văn trên thế giới.

Xác định khuôn mặt người trong ảnh là một bài toán hấp dẫn, đã thách thức nhiều người nghiên cứu vì tính ứng dụng to lớn trong thực tế. Hy vọng với bài viết này sẽ giúp nhiều người có được những kiến thức nhất định, giúp đỡ nhiều người không phải mất nhiều thời gian khi bắt đầu nghiên cứu bài toán hấp dẫn này, để cùng nhau vươn ra biển khơi tri thức của thế giới.





**Hình 16:** Các khó khăn của việc xác định mặt người:  
 (a) hướng mặt nghiêng; (b) mắt kính đen và nón; (c) ảnh bị chói bởi đèn;  
 (d) máy ảnh đặt phía trên và sau lưng người bị chụp;  
 (e) vùng da các khuôn mặt dính nhau;  
 (f) màu ở môi trường xung quanh gần với màu da người;  
 (g) chất lượng ảnh kém.

## TÀI LIỆU THAM KHẢO

- [1] Rein-Lien Hsu, Mohamed abdel-Mottaleb, and Anil K. Jain, "Face Detection in Color Images", IEEE Transaction on Pattern Analysis and Machine Intelligent, vol. 24, no. 5, pp. 696-706, 2002.
- [2] C. Garcia, G. Zikos, and G. Tziritas, "Face Detection in Color Images using Wavelet Packet Analysis", Proc. of IEEE International Conference on Multimedia Computing and System, vol. 1, pp. 703-708, IEEE, 1999.
- [3] Saman Cooray and Noel O'Connor, "Facial Feature Extraction and Principal Component Analysis for Face Detection in Color Image", ICIAR 2004, LNCS 3212, pp. 741-749, Springer-Verlag Berlin Heidelberg, 2004.
- [4] Satyanadh, Li Tao, and Vijayan Asari, "Face Detection Technique Based on Intensity and Skin Color Distribution", International Conference on Image Processing, IEEE, 2004.
- [5] Rainer Stiefelhausen, Jie Yang, and Alex Waibel, "Tracking Eyes and monitoring Eye Gaze", Proceedings of the Workshop on Perceptual User Interfaces (PUI'97), Alberta, Canada. pp. 98-100, 1997.
- [6] Carlos Morimoto, Dave Koons, Amon Amir, and Myron Flickner, "Real-Time Detection of Eyes and Faces", In Proc. Workshop on Perceptual User Interfaces, pp. 117-120, 1998.
- [7] Eun Yi Kim, Sin Kuk Kang, Keechul Jung, and Hang Joon Kim, "Eye Mouse: Mouse Implementation using Eye Tracking", IEEE, 2005.
- [8] Hichem Sabbi and Nozha Boujemaa, "Coarse to Fine Face Detection Based on Skin Color Adaption", Biometric Authentication, LNCS 2359, pp. 112-120, Springer-Verlag Berlin Heidelberg, 2002.
- [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] Douglas Chai and King N. Ngan, "Face Segmentation Using Skin-Color map in Videophone Applications", IEEE Transaction on Circuits and Systems for Video Technology, vol. 9, no. 4, 1999.
- [11] Christophe Garcia and Georgios Tzirita, "Face Detection Using Quantized Skin Color Region Merging and Wavelet Packet Analysis", IEEE Transaction on Multimedia, vol. 1, no. 3, 1999.
- [12] Mark Everingham and Andrew Zisserman, "Automated Person Identification in Video", CIVR 2004, LNCS 3115, pp. 289-298, Springer-Verlag Berlin Heidelberg, 2004.
- [13] Fuzhen Huang and Jianbo Su, "Multiple Face Contour Detection Using adaptive Flows", Sinobiometrics 2004, LNCS 3338, pp. 137-143, Springer-Verlag Berlin Heidelberg, 2004.
- [14] Seonghoon Kang and Seong-Wan Lee, "Object Detection and Classification for Outdoor Walking Guidance System", BMCV 2002, LNCS 2525, pp. 601-610, Springer-Verlag Berlin Heidelberg, 2002.
- [15] Shihong Lao and Masato Kawade, "Vision-Based Face Understanding Technologies and Their Application", Sinobiometric 2004, LNCS 3338, pp. 339-348, Springer-Verlag Berlin Heidelberg, 2004.
- [16] Stephen C. Y. Chan and Paul H. Lewis, "A Pre-filter Enabling Fast Frontal Face Detection", Visual'99, LNCS 1614, pp. 777-785, Springer-Verlag Berlin Heidelberg, 1999.
- [17] Klaus J. Kirehberg, Oliver Jeorsky and Robert W. Frischholz, "Genetic Model Optimization for Hausdorff Distance-Based Face Localization", Biometric Authentication, LNCS 2359, pp. 103-111, Springer-Verlag Berlin Heidelberg, 2002.
- [18] Stanley M. Bileschi and Bernd Heisele, "Advances in Component-Based Face Detection", SVM 2002, LNCS 2388, pp. 135-143, Springer-Verlag Berlin Heidelberg, 2002.
- [19] Rainer Lienhart, Alexander Kuranov, and Vadim Pisarevsky, "Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection", DAGM 2003, LNCS 2781, pp. 297-304, Springer-Verlag Berlin Heidelberg, 2003.
- [20] Jean-Christophe Terrillon, Mahdad N. Shirazi, Daniel McReynolds, Mohamed Sadek, Yunlong Sheng, Shigeru Akamatsu, and Kazuhiko Yamamoto, "Invariant Face Detection in Color Image Using Orthogonal Fourier-Mellin Moments and Support Vector Machines", ICAPR 2001, LNCS

- 2013, pp. 83-92, Springer-Verlag Berlin Heidelberg, 2001.
- [21] Hannes Kruppa, Martin A. Bauer, and Bernt Schiele, "Skin Patch Detection in Real-World Images", DAGM 2002, LNCS 2449, pp. 109-116, Springer-Verlag Berlin Heidelberg, 2002.
- [22] Jia Kui and Liyanage C. De Silva, "Combined Face Detection/Recognition System for Smart Rooms", AVBPA 2003, LNCS 2688, pp. 787-795, Springer-Verlag Berlin Heidelberg, 2003.
- [23] Jin Ok Kim, Sung Jin Seo, Chin Hyun Chung, Jun Hwang, and Woonglae Lee, "Face Detection by Facial Features with Color Images and Face Recognition Using PCA", ICCSA 2004, LNCS 3043, pp. 1-8, Springer-Verlag Berlin Heidelberg, 2004.
- [24] YoungOul Kim, SungHo Jang, SangJin Kim, chang-Woo Park, and Joonki Paik, "Pose-Invariant Face Detection Using Edge-Like Blob Map and Fuzzy Logic", IEA/AIE 2005, LNCS 3533, pp. 695-704, Springer-Verlag Berlin Heidelberg, 2005.
- [25] Bernhard Froba and Walter Zink, "On the Combination of Different Template Matching Strategies for Fast Face Detection", MCS 2001, LNCS 2096, pp. 418-428, Springer-Verlag Berlin Heidelberg, 2001.
- [26] Henry Schneiderman, "Statistical Approach to 3D Object Detection – Applied to Faces and Cars", Ph.D. Thesis, Carnegie Mellon University, 2000.
- [27] Kang Ryoung Park, "Gaze Detection System by Wide and Auto Pan/Tilt Narrow View Camera", DAGM 2003, LNCS 2781, pp. 76-83, Springer-Verlag Berlin Heidelberg, 2003.
- [28] Michael Elad, Yacov Hel-Or, and Renato Keshet, "Pattern Detection Using Maximal Rejection Classifier", LNCS 2059, Springer-Verlag Berlin Heidelberg, 2001.
- [29] Gil Friedrich and Yehezkel Yeshurun, "Seeing People in the Dark: Face Recognition in Infrared Images", BMVC 2002, LNCS 2525, pp. 348-359, Springer-Verlag Berlin Heidelberg, 2002.
- [30] Hong-Mo Je, Daijin Kim, and Sung Yang Bang, "Human Face Detection in Digital Video Using SVM Ensemble", Neural Processing Letter 17:239-252, Kluwer Academic Publishers, Netherlands, 2003.
- [31] Jongmoo Choi, Sanghoon Lee, Chilgee Lee, and Juneho Yi, "PrimeEye: A Real-Time Face Detection and Recognition System Robust to Illumination Changes", AVBPA 2001, LNCS 2091, pp. 360-365, Springer-Verlag Berlin Heidelberg, 2001.
- [32] A Jonathan Howell and Hilary Buxton, "Visual Mediated Interaction Using Learnt Gestures and Camera Control", GW 2001, LNAI 2298, pp. 272-284, Springer-Verlag Berlin Heidelberg, 2002.
- [33] Daniel Keren, Margarita Osadchy, and Craig Gotsman, "Antifaces: A Novel, Fast Method for Image Detection", IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 23, no. 7, IEEE, 2001.
- [34] Gines Garcia Mateos and Cristina Vicente Chicote, "Face Detection on Still Images Using HIT Maps", AVBPA 2001, LNCS 2091, pp. 102-107, Springer-Verlag Berlin Heidelberg, 2001.
- [35] M. Castrillon Santana, J. Lorenzo Navarro, J. Cabrera Gamez, F. M. Hernandez Tejera, and J. Mendez Rodriguez, "Detection of Frontal Faces in Video Streams", Biometric Authentication, LNCS 2359, pp. 91-102, Springer-Verlag Berlin Heidelberg, 2002.
- [36] Chang-Woo Lee, Yeon-Chul Lee, Sang-Yong Bak, and Hang-Joon Kim, "Real-Time Face Detection and Tracking Using PCA and NN", PRICAI 2002, LNAI 2417, pp. 615, Springer-Verlag Berlin Heidelberg, 2002.
- [37] Jie Chen, Yuemin Li, Laiyun Qing, Baocai Yin, and Wen Gao, "Face Samples Re-lighting for Detection Based on the Harmonic Images", PCM 2004, LNCS 3332, pp. 585-592, Springer-Verlag Berlin Heidelberg, 2004.
- [38] Hazem El-Bakry, "A Rotation Invariant Algorithm for Recognition", Fuzzy day 2001, LNCS 2206, pp. 284-290, Springer-Verlag Berlin Heidelberg, 2001.
- [39] Kenji Iwata, Hitoshi Hongo, Kazuhiko Yamamoto, and Yoshinori Niwa, "Robust Facial Parts Detection by Using Four Directional Features and Relaxation Matching", KES 2003, LNAI 2774, pp. 882-888, Springer-Verlag Berlin Heidelberg, 2003.
- [40] Ai Haizhou, Liang Luhong, and Xu Guangyou, "A General Framework for Face Detection", ICMI 2000, LNCS 1948, pp. 119-126, Springer-Verlag Berlin Heidelberg, 2000.
- [41] Hongmin Zhang, Debin Zhao, Wen Gao, and Xilin Chen, "Combining Skin Color Model and Neural Network for Rotation Invariant Face Detection", ICMI 2000, LNCS 1948, pp. 237-244, Springer-Verlag Berlin Heidelberg, 2000.
- [42] Rob McCready, "Real-Time Face Detection on a Configurable Hardware System", FPL 2000, LNCS 1896, pp. 157-162, Springer-Verlag Berlin Heidelberg, 2000.
- [43] J. Fritsch, S. Lang, M. Kleinhagenbrock, G. A. Fink, and G. Sagerer, "Improving Adaptive Skin Color Segmentation by Incorporating Results from Face Detection", In Proc. IEEE Int. Workshop on Robot and Human Interactive Communication, Germany, IEEE, 2002.
- [44] Raquel Montes Diez, Cristina Conde, and Enrique Cabello, "Automatic Detection of the Optimal Acceptance Threshold in a Face Verification System", BioAW 2004, LNCS 3087, pp. 70-79, Springer-Verlag Berlin Heidelberg, 2004.
- [45] Henry Nicponski, "Understanding In-Plane Face Rotations Using Integral Projections", ICIAR 2004, LNCS 3213, pp. 633-642, Springer-Verlag Berlin Heidelberg, 2004.

- [46] Tony W.H. Ao leong and Raymond S.T. Lee, "iJade Face Recognizer – A Multi-agent Based Pose and Scale Invariant Human Face Reconition System", KES 2004, LNAI 3214, pp. 594-601, Springer-Verlag Berlin Heidelberg, 2004.
- [47] Jin Ok Kim, Jin Soo Kim, and Chin Hyun Chung, "Face Region Dectection on Skin Chrominance from Color Images by Facial Features", ICADL 2004, LNCS 3334, pp. 646, Springer-Verlag Berlin Heidelberg, 2004.
- [48] Henry A. Rowley, Shumeet Baluja, and Takeo Kanade, "Neural Network-Based Face Detection", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 1, pp. 23-38, IEEE, 1998.
- [49] Henry A. Rowley, Shumeet Baluja, and Takeo Kanade, "Rotation Invariant Neural Network-Based Face Detection", Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 38-44, 1998
- [50] Stephanie Jehan-Besson, Michel Barlaud, and Gilles Aubert, "DREAMS: Deformable Region Driven by an Eulerian Accurate Minimization Method for Image and Video Segmentation (Application to Face Detection in Color Video Sequences)", ECCV 2002, LNCS 2352, pp. 365-380, Springer-Verlag Berlin Heidelberg, 2002.
- [51] Hongming Zhang and Debin Zhao, "Spatial Histogram Features for Face Detection in Color Images", PCM 2004, LNCS 3331, pp. 377-384, Springer-Verlag Berlin Heidelberg, 2004.
- [52] Paul Viola, "Robust Real-Time Face Detection", International Journal of Computer Vision 57(2), 137-154, Kluwer Academic Publishers, Netherlands, 2004.
- [53] Kim-Fung Jang, Ho-Man Tang, Michael R. Lyu, and Irwin King, "A Face Processing System Based on Committee Machine: The Approach and Experimental Results", CAIP 2003, LNCS 2756, pp. 614-622, Springer-Verlag Berlin Heidelberg, 2003.
- [54] James W. Davis and Serge Vals, "A Perceptual User Interface for Recognizing Head Gesture Acknowledgements", ACM International Conference Proceeding Series, vol. 15, pp. 1-7, 2001.
- [55] Chengjun Liu and Harry Wechsler, "Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition", IEEE Transactions on Iamge Processing, vol. 11, no. 4, IEEE, 2002.
- [56] Wenchao Zhang, Shiguang, Wen Gao, Xilin Chen, and Hongming Zhang, "Local Gabor Binary Pattern Histogram Sequence (LGBPHS): A Novel Non-Statistical Model for Face Representation and Recognition", Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV'05), IEEE, 2005.
- [57] Peter N. Belhumeur, "Ongoing Challenges in Face Recognition", Frontiers of Engineering: Reports on Leading-Edge Engineering from the 2005, Symposium (2006).
- [58] Juwei Lu, K.N. Plataniotis, A.N. Venetsanopoulos, and Stan Z. Li, "Ensenble-based Discriminant Learning with Boosting For Face Recognition", IEEE Transactions on Neural Networks, vol. 17, pp. 166-178, 2006.
- [59] Rogerio S. Feris, Jim Gemmell, Kentaro Toyama, and Volker Kruger, "Hierarchical Wavelet Networks for Facial Feature Localization", Proceedings of Fifth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 118-123, 2002.
- [60] Baochang Zhang, "Gabor-Kernel Fisher Analysis for Face Recognition", LNCS 3332, Springer-Verlag Berlin Heidelberg, 2004.
- [61] Paola Campadelli, Raffaella Lanzarotti, Chiara Savazzi, "A feature-based face recognition system", Proceedings of 12th International Conference on Image Analysis and Processing, pp. 68-73, IEEE, 2003.
- [62] Julien Meynet, Vlad Popovici, and Jean-Philippe Thiran, "Face Detection with Mixtures of Boosted Discriminant Features", Technical Report, [http://biomed.epfl.ch/index.php?module=people&PEOPLE\\_MAN\\_OP=view&PHPWS\\_MAN\\_ITEMS%5B%5D=14&show\\_publications=1](http://biomed.epfl.ch/index.php?module=people&PEOPLE_MAN_OP=view&PHPWS_MAN_ITEMS%5B%5D=14&show_publications=1), 2005.
- [63] Filareti Tsalakanidou, Sotiris Malassiotis, and Michael G. Strintzis, "Face Localization and Authentication Using Color and Depth Images", IEEE Transactions on Image Processing, vol. 14, no. 2, IEEE, 2005.
- [64] Yasunori Ishii, Kazuyuki Imagawa, Eiji Fukumiya, Katsuhiro Iwasa, Yasunobu Ogura, "Profile Face Detection using Block Difference Feature For Automatic Image Annotation", International Conference on Consumer Electronics ICCE '06. 2006 Digest of Technical Papers, IEEE, 2006.
- [65] Jong-Bae Kim, Su-Woong Jung, and Hang-Joon Kim, "Face Detection by Integrating Multiresolution-Based Watersheds and Skin-Color Model", IEA/AIE 2002, LNAI 2358, pp. 715-724, Springer-Verlag Berlin Heidelberg, 2002.
- [66] Gary G. Yen and Nethrie Nithianandan, "Automatic Facial Feature Extraction Using Edge Distribution and Genetic Search", International Journal of Computational Intelligence and Applications, vol. 3, no. 1, pp. 89-100, Imperial College Press, 2003.
- [67] Hideaki Sato, Katsuhiro Sakamoto, Yasue Mitsukura, and Norio Akamatsu, "Face Edge Detection System by Using the GAs", KES 2004, LNAI 3213, pp. 847-852, Springer-Verlag Berlin Heidelberg, 2004.
- [68] J.S. Taur and C.W. Tao, "A New Neuro-Fuzzy Classifier with Application to On-Line Face Detection and Recognition", Journal of VLSI Signal Processing 26, 397-409, Kluwer Academic Publishers, the Netherlands, 2000.
- [69] Sang-Burm Rhee and Young-Hwan Lee, "Improved Face Detection Algorithm in Mobile Enviroment", ICCS 2004, LNCS 3036, pp. 638-686, Springer-Verlag Berlin Heidelberg, 2004.

- [70] Douglas Chai and Kim N. Ngan, "Locating Facial Region of a Head-and-Shoulders Color Image", Proc. Third Int'l Conf. Automatic Face and Gesture Recognition, pp. 124-129, 1998.
- [71] Bae-Ho Lee, Kwang-Hee Kim, Yonggwon Won, and Jiseung Nam, "Efficient and Automatic Faces Detection Based on Skin-Tone and Neural Network Model", IEA/AIE 2002, LNAI 2358, pp. 57-66, Springer-Verlag Berlin Heidelberg, 2002.
- [72] Shu-Fai Wong and Kwan-Yee Keneth Wong, "Fast Face Detection Using QuadTree Based Color Analysis and Support Vector Verification", ICIAR 2004, LNCS 3212, pp. 676-683, Springer-Verlag Berlin Heidelberg, 2004.
- [73] Henry Schneiderman and Takeo Kanade, "Probabilistic Modeling of Local Appearance and Spatial Relationship for Object Recognition", Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 45-51, IEEE, 1998.
- [74] Gines Garcia Mateos, Alberto Ruiz, and Perdo E. Lopez-de-Teruel, "Face Detection Using Integral Projection Models", SSPR&SPR 2002, LNCS 2396, pp. 644-653, Springer-Verlag Berlin Heidelberg, 2002.
- [75] QingHua Wang, Luis Seabra Lopes, and David M.J. Tax, "Visual object recognition through one-class learning", Image analysis and recognition: international conference ICIAR, vol. 1, pp. 463-470, 2004.
- [76] Selin Baskan, M. Mete Bulut, Volkan Atalay, "Projection based method for segmentation of human face and its evaluation", Pattern Recognition Letters 23, pp. 1623-1629, Elsevier Science, 2002.
- [77] Thang V. Pham. Marcel Worring, and Arnold W.M. Smeulders, "Face Detection by Aggregated Bayesian Network Classifiers", MLDM 2001, LNAI 2123, pp. 249-262, Springer-Verlag Berlin Heidelberg, 2001.
- [78] Shou-Der Wei and Shang-Hong Lai, "An Efficient Algorithm for Detecting Faces from Color Images", PCM 2002, LNCS 2532, pp. 1177-1184, Springer-Verlag Berlin Heidelberg, 2002.
- [79] Yen-Yu Lin, Tyng-Luh Liu, and Chiou-Shann Fuh, "Fast Object Detection with Occlusions", ECCV 2004, LNCS 3021, pp. 402-413, Springer-Verlag Berlin Heidelberg, 2004.
- [80] Kyongpil Min, Junchul Chun, and Goorack Prak, "A Nonparametric Skin Color Model for Face Detection from Color Images", PDCAT 2004, LNCS 3320, pp. 116-120, Springer-Verlag Berlin Heidelberg, 2004.
- [81] Herve Abdi, "A Generalized Approach for Connectionist Auto-Associative Memories: Interpretation, Implication & Illustration for Face Processing", Artificial intelligence and cognitive sciences, Manchester University Press. pp. 149-165, Manchester, 1988.
- [82] Jianping Fan, David K.Y. Yau, Ahmed K. Elmagarmid, an Walid G. Arref, IEEE Transactions on Image Processing, vol. 10, no. 10, IEEE, 2001.
- [83] Jianzhong Fang and Guoping Qiu, "Learning an Information Theoretic Transform for Object Detection", ICAR 2004, LNCS 3211, pp. 503-510, Springer-Verlag Berlin Heidelberg, 2004.
- [84] T. Darrell, G. Gordon, M. Harville, and J. Woodfill, "Integrated Person Tracking Using Stereo, Color, and pattern Detection", International Journal of Computer Vision 37(2), 175-185, Kluwer Academic Publishers, the Netherlands, 2000.