TẠP CHÍ KHOA HỌC & CÔNG NGHỆ ĐẠI HỌC DUY TÂN
DTU Journal of Science and Technology 01(32) (2019) 54-63

TRƯỜNG ĐẠI HỌC DUY TÂN
DUYTAN UNIVERSITY

Một hướng tiếp cận nâng cao hiệu quả phát hiện mặt người trong ảnh

An effective approach to face detection

Nguyễn Thị Bích Hạnh[a,*], Trần Lê Thăng Đồng[b], Trương Văn Trương[a]
Bich Hanh Nguyen, Thang Dong Tran Le, Van Truong Truong

[a]*Khoa Điện Điện tử, Đại học Duy Tân, 03 Quang Trung, Đà Nẵng, Việt Nam*
Faculty of Electrical & Electronics Engineering, Duy Tan University, 03 Quang Trung, Danang, Vietnam
[b]*Trung tâm Điện - Điện tử, Đại học Duy Tân, 03 Quang Trung, Đà Nẵng, Việt Nam*
Center of Electrical Engineering, Duy Tan University, 03 Quang Trung, Danang, Vietnam

(Ngày nhận bài: 07/12/2018, ngày phản biện xong: 21/01/2019, ngày chấp nhận đăng: 16/02/2019)

Tóm tắt

Phát hiện khuôn mặt là một trong những mắc xích quan trọng trong quy trình nhận dạng mặt người. Đã có nhiều hướng tiếp cận phát hiện khuôn mặt đã và đang được nghiên cứu, theo đó bài báo này đề xuất một hướng tiếp cận phát hiện mặt người theo cấu trúc khuôn mặt. Chúng tôi sử dụng đặc trưng Ridge và Valley để tìm ra các điểm lồi lõm trên khuôn mặt như mắt, mũi, miệng làm cơ sở cho việc xác định khuôn mặt. Các kết quả thực nghiệm so với các phương pháp khác cho thấy đây là hướng tiếp cận hiệu quả với độ chính xác cao, có thể ứng dụng trong các ứng dụng phát hiện thời gian thực, đặt biệt là trong các ứng dụng phát hiện cảnh báo ngủ gật trên xe ô tô.

Từ khóa: Phát hiện mặt người; Ridge and Valley; tăng cường; Haar-like;

Abstract

Face detection is one of the important phases in the face recognition process. There have been many approaches to face detection that have been studied. In this paper, we propose an approach to detect human face according to face structure. We use Ridge and Valley features to find the convex and concave points on the face such as eyes, nose, and mouth as a basis for facial identification. The experimental results compared with other methods show that this is an effective approach with high accuracy, and can be applied in real-time detection applications, especially in system of drowsiness on cars.

Keywords: Face detection; Ridge and Valley; Boosting; Haar-like.

1. Giới thiệu

Bài toán phát hiện và nhận dạng đối tượng là một trong những vấn đề cơ bản thuộc lĩnh vực thị giác máy. Phát hiện và nhận dạng mặt người trong dữ liệu ảnh và video là bài toán có nhiều ứng dụng thực tiễn như hệ thống giám sát an ninh, hệ thống giao tiếp người máy, hỗ trợ lái xe an toàn... Là một trong những mắc xích quan trọng trong hệ thống nhận dạng mặt người, bài toán phát hiện khuôn mặt đã và đang dành được sự quan tâm nghiên cứu của nhiều nhà nghiên cứu trong suốt nhiều thập kỷ qua. Tuy nhiên, bài toán phát hiện khuôn mặt vẫn gặp nhiều khó khăn do tư thế, góc chụp, thiếu hoặc bị che một số thành phần của khuôn mặt... nên cho đến nay các nghiên cứu vẫn chưa đạt được kết quả như mong muốn.

Email: hanhnguyen.unilab@gmail.com

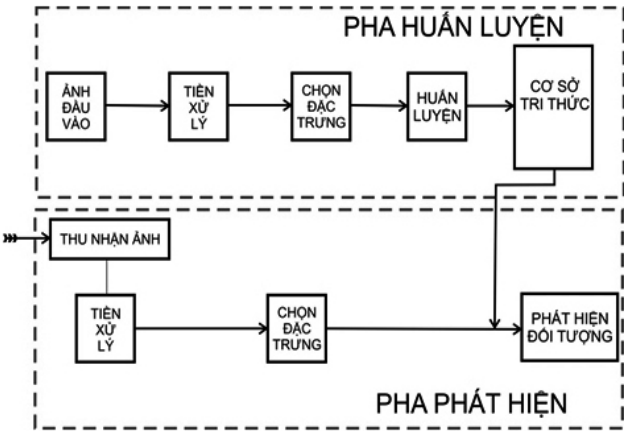
Đã có nhiều hướng tiếp cận được đề xuất đối với bài toán phát hiện khuôn mặt: dựa trên các đặc trưng bất biến của khuôn mặt [1], mối quan hệ giữa các thành phần trên khuôn mặt [2], mô hình hỗn hợp Gaussian hỗn hợp [7], mô hình Markov ẩn (HMM) [10], mạng nơ-ron nhân tạo [9], máy véc tơ tựa (SVM) [8], các mô hình xác suất [11]... hoặc gần đây là Hong Du [6] với kỹ thuật phân tích Ridge và Valley hiểu theo nghĩa thật sự của nó là lồi (chỏm) là các điểm cao nhất trên dãy núi và lõm (thung lũng) là các điểm có độ cao thấp nhất, các điểm này dùng để biểu diễn được các vùng như: mắt, chân mày, mũi, và miệng. Tuy nhiên, Canton-Ferrer [12] tuyên bố rằng điều kiện ánh sáng, môi trường, các góc camera hướng mặt về phía máy sẽ ảnh hưởng đáng kể đến hiệu suất của các mô hình. Do đó chúng tôi đề xuất cải tiến phương pháp của Hong Du bằng cách rút trích các điểm Ridge và Valley và mô hình hóa chúng thành các khung lưới mô hình hóa 3D mô phỏng khuôn mặt từ đó cải thiện hiệu xuất phát hiện khuôn mặt.

Bài báo này được cấu trúc như sau: Phần 2 sẽ giới thiệu các kết quả và các công trình có liên quan đến bài toán phát hiện khuôn mặt, hướng tiếp cận đặc trưng Ridge và Valley. Phần 3 đề xuất hướng tiếp cận mô hình hóa 3D cho việc phát hiện đối tượng dựa trên đặc trưng Ridge và Valley. Phần 4 trình bày các kết quả thực nghiệm và so sánh hiệu quả giữa các hướng tiếp cận. Cuối cùng là phần kết luận và hướng phát triển.

2. Các kết quả và công trình liên quan

2.1. Hướng tiếp cận học máy

Học máy là một trong các hướng tiếp cận hiệu quả đối với bài toán phát hiện và nhận dạng đối tượng. Mô hình học máy có 2 pha: pha huấn luyện (training) và pha phát hiện (testing). Pha huấn luyện được thực hiện trước để tạo ra cơ sở tri thức phục vụ cho pha testing (Hình 1).



Hình 1. Phát hiện đối tượng dựa trên mô hình học máy

A. Tiền xử lý

Ảnh đầu vào có thể bị nhiễu do nhiều nguyên nhân. Vì vậy cần nâng cao chất lượng ảnh để tăng độ chính xác cho hệ thống. Chức năng chính của bộ tiền xử lý là lọc nhiễu, nâng độ tương phản để làm ảnh rõ hơn, nét hơn [15]...

B. Lựa chọn đặc trưng

Có nhiều phương pháp lựa chọn đặc trưng cho bài toán phát hiện đối tượng. Một trong những đặc trưng hiệu quả được nhiều nhà nghiên cứu quan tâm, đó là đặc trưng Haar Like, sử dụng đặc trưng này có thể đảm bảo yêu cầu đáp ứng thời gian thực cho hệ thống. Đặc trưng Haar-like do Viola và Jones đề xuất [4], gồm 4 đặc trưng cơ bản. Mỗi đặc trưng Haar-like là sự kết hợp của hai hay ba hình chữ nhật "trắng" hay "đen". Đặc trưng Haar-like là sự chênh lệch giữa tổng của các pixel của các vùng đen và các vùng trắng như trong công thức sau:

$$f(x) = \sum_{Vùngđen} pixels - \sum_{Vùngtrắng} pixels$$

Để tính các giá trị của đặc trưng Haar-like, cần phải tính tổng của các điểm ảnh trên ảnh. Nhưng việc tính các giá trị của các đặc trưng Haar-like cho tất cả các vị trí trên ảnh đòi hỏi chi phí tính toán khá lớn, không đáp ứng được cho các ứng dụng thời gian thực. Do đó Viola và Jones đưa ra một khái niệm gọi là "ảnh tích hợp" (Integral Image) [4]. Ảnh tích hợp có các tính chất sau: giá

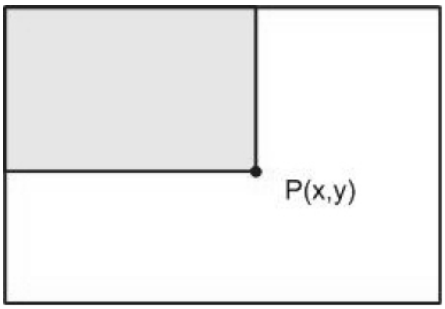
trị pixel tại vị trí (x, y) bất kỳ trong ảnh tích hợp bằng tổng số các pixel thuộc hình chữ nhật được tạo bởi góc trái trên cùng của ảnh (0, 0) và điểm đang xét.

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

Trong đó:

i: ảnh ban đầu.

P: ảnh tích hợp của i.



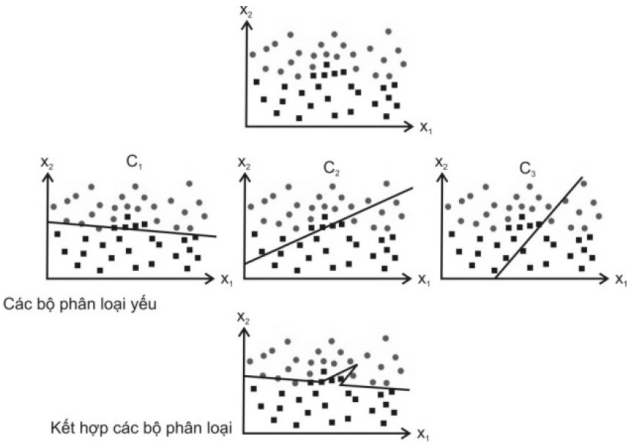
Hình 2. Ảnh chia nhỏ tại tọa độ (x, y)

Việc tính ảnh tích hợp sẽ được thực hiện theo kỹ thuật quy hoạch động.

C. Huấn luyện

Các thuật toán học máy thường áp dụng kỹ thuật phân tích, thống kê để xấp xỉ với một hàm phân lớp tuyến tính. Có nhiều mô hình học máy đã được đề xuất cho bài toán phát hiện khuôn mặt: Mô hình mạng Nơ-ron do H. Rowley và cộng sự đề xuất năm 1998 [9], phương pháp máy véctơ tựa (SVM - Support Vector Machine do nhóm E. Osuna đề xuất năm 1997 [8], mô hình phân lớp Bayes doH. Schneiderman và T. Kanade đề xuất năm 1998 [11], mô hình Markov ẩn do A. Rajagopalan và các cộng sự đề xuất năm 1998 [10]. Đặc biệt là mô hình học tăng cường (Boosting) do P. Viola và M. Jones đề xuất năm 2001 [4] và mô hình FloatBoost do StanZ. Li và Zhen Qiu Zhang 2004 [12].

Nguyên lý cơ bản của học tăng cường là tổ hợp tuyến tính của các máy phân loại yếu để hình thành một máy phân loại mạnh (Hình 3). Với phương pháp này, một máy phân loại yếu (với tỉ lệ thành công nhỏ hơn 50%) có thể được nâng cấp ("boost") thành một máy phân loại mạnh với mục tiêu là lỗi huấn luyện của máy phân loại mới sẽ tiệm cận đến Zero với thời gian thực hiện đa thức.



Hình 3. Kết hợp các bộ phân loại yếu thành bộ phân loại mạnh

D. Cơ sở tri thức

Sau khi huấn luyện xong sẽ có được một "cơ sở tri thức" (Knowledge base)để phục vụ cho pha testing tìm kiếm và phát hiện đối tượng. Đây là yếu tố xương sống của toàn bộ hệ thống. Nó được tạo ra ở bước huấn luyện và được lưu trữ dưới dạng file dữ liệu có cấu trúc.

E. Phát hiện đối tượng (Pha testing)

Pha phát hiện đối tượng được thực hiện dựa vào cơ sở tri thức thu được sau quá trình huấn luyện.

2.2. Hướng tiếp cận đặc trưng Ridge và Valley

Hướng tiếp cận này quan tâm đến các đặc trưng lồi lõm của các đối tượng trong ảnh. Như vậy các đối tượng trong ảnh sẽ được trích chọn và biểu diễn dưới dạng khung lưới 3D. Mô hình khung lưới được kiến tạo bằng cách xác định các cạnh của vật thể, hoặc bằng cách nối các điểm đỉnh hợp thành của một đối tượng dùng đường thẳng hoặc cong. Đối với đặc trưng khuôn mặt, chỉ cần sử dụng các đối tượng điểm và đường thẳng.

Liên quan tới việc biểu diễn chiều thứ ba bằng độ sâu của màu, James L. Crowley đã đề xuất các khái niệm *Ridge* và *peak* năm 1983 [1]. Trong đó, *Ridge* là các điểm lồi trên ảnh, tập các điểm *Ridge* sẽ tạo thành những đường xương sống và các đường này sẽ tạo thành các chùm tia gọi là *peak*. James L. Crowley đã sử dụng phép hiệu của lọc thông thấp để rút ra các điểm lồi trên ảnh, sau đó sử dụng một thuật toán duyệt để kết nối chúng lại với nhau thành các đặc trưng *Ridge* và *peak*.

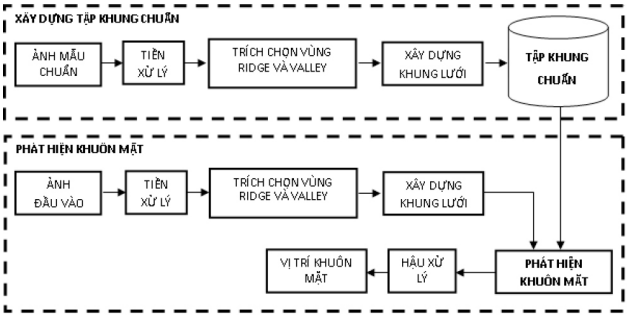
Quan Yuan đã áp dụng các loại đặc trưng này cho bài toán phát hiện khuôn mặt [13] và đưa ra khái niệm đặc trưng *tựa thung lũng hay* đặc trưng *có dạng rãnh* (*Valley-like feature*). Đặc trưng này dựa trên hiệu mức sáng của các điểm xung quanh và điểm đang xét ở trung tâm để xác định xem điểm này có dạng rãnh hay không. Hướng tiếp cận này giúp nhận biết được các vị trí trên ảnh có đặc điểm của mắt, mũi, miệng. Liên quan đến đặc trưng *Ridge*, Hai Tran và A.Lux [3] đã đề xuất phương pháp sử dụng bộ lọc Laplacian để tăng cường thông tin cạnh sau đó dò tìm *Ridge* và *peak* dưới nhiều mức khác nhau. Trong hướng tiếp cận này, các đặc trưng *Ridge* và *peak* được biểu diễn dưới dạng các đồ thị quan hệ hoặc cây phân cấp với các tầng là các *Ridge* và *peak* được dò tìm tại các mức khác nhau.

Gần đây nhất, Hong Du [6] xây dựng cây cấu trúc *Ridge* và *Valley* để phát hiện mặt người trong ảnh kỹ thuật số. Theo cách tiếp cận này Hong Du cấu trúc lại khuôn mặt theo cây phân cấp của các điểm *Ridge* và *Valley*. Các điểm lồi lõm được phát hiện ở nhiều mức khác nhau để dò tìm chi tiết các vùng trên khuôn mặt từ mức cao đến mức thấp nhất.

Với các hướng tiếp cận trên, ảnh khuôn mặt trong ảnh 2D có thể biểu diễn được bằng phương pháp mô hình hóa theo dạng 3D.

3. Đề xuất mô hình phát hiện khuôn mặt

Khuôn mặt sẽ được biểu diễn dựa trên mô hình khung lưới theo một cấu trúc bao gồm các cạnh liên kết với nhau qua các đỉnh được tạo thành bởi các vùng *Ridge* và *Valley*. Mỗi nút trong khung lưới sẽ có thể là nút *Ridge* hoặc nút *Valley*, điều này sẽ giúp nhận biết được độ cao hoặc thấp của các vị trí trên khung lưới. Dựa vào mô hình khung lưới này để dò tìm và phát hiện các khuôn mặt có trên ảnh với nhiều mức khác nhau.

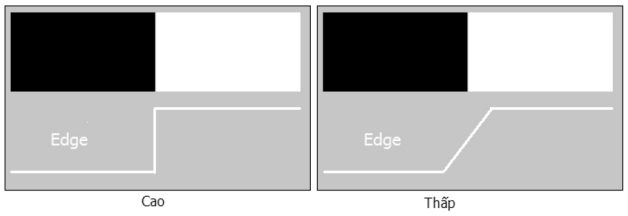


Hình 4. Mô hình phát hiện khuôn mặt dựa trên đặc trưng Ridge và Valley

3.1. Tiền xử lý

Ảnh đầu vào sẽ được chuyển sang ảnh đa cấp xám và lọc nhiễu. Trong phần này chúng tôi sử dụng kết hợp bộ lọc Trung bình (mean filter) và bộ lọc Gaussian vì hai bộ lọc này không làm mờ các cạnh.

Tùy chất lượng ảnh đầu vào, ảnh đầu ra nhận được sau khi lọc nhiễu sẽ được làm mịn tương đối, từ đó ảnh cũng có độ nét khác nhau.



Hình 5. Độ nét của ảnh

Hình 5 thể hiện tốc độ chuyển tiếp thông tin của hình ảnh tại biên, vì vậy độ nét cao cho kết quả là các chuyển tiếp hình dạng và chi tiết biên được thể hiện một cách rõ ràng. Độ nét của ảnh có thể nhận dạng qua các đường biên trong ảnh, và điều này sẽ ảnh hưởng tới việc trích chọn các vùng *Ridge* và *Valley*. Do đó để tăng cường độ nét của ảnh và làm nổi bật các đường viền trong ảnh, chúng tôi áp dụng hai thuật toán cơ bản là phép giãn (Dilation) và phép co (Erosion) trong phương

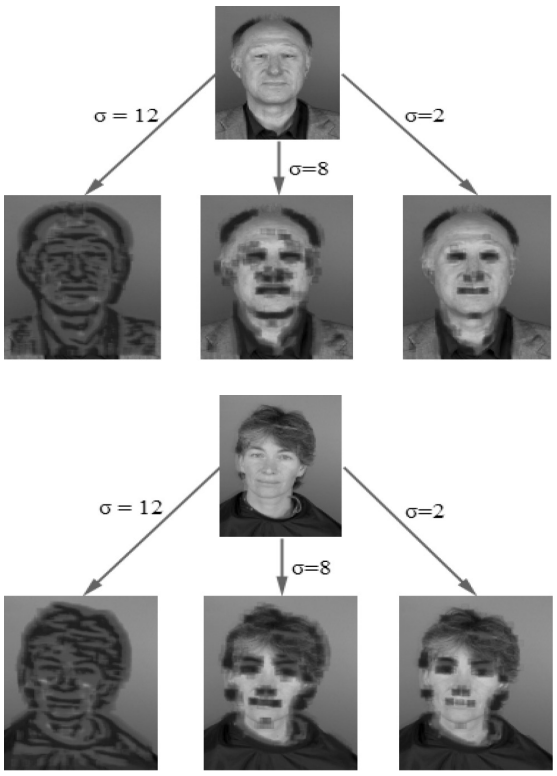
pháp hình thái. Bước xử lý này sẽ làm rõ một số phần trong ảnh và một số phần khác sẽ mờ đi.

3.2. Trích chọn đặc trưng Ridge và Valley

Trong ảnh khuôn mặt, mũi và các phần nhô của hai má được xem là *Ridge* còn các phần hốc mắt, lỗ mũi, miệng được được xem là *Valley*. Các ảnh sau bước tiền xử lý sẽ được phân tích để dò tìm quỹ tích của các điểm cực trị. Hình 6 và Hình 7 biểu diễn các đặc trưng *Ridge* và *Valley* tương tự như các kết quả của HaiTran [3].

Hình 6. Tìm thông tin Ridge và Valley trên khuôn mặt sử dụng LoG với $\sigma = \sqrt{2}$.

Các điểm *Ridge* được tô màu đỏ và *Valley* được tô màu xanh. Với kết quả này, chúng ta lưu ý rằng các điểm *Ridge* và *Valley* rút trích được chỉ biểu diễn được các vùng *Ridge* và *Valley* nhỏ và chi tiết, chẳng hạn như: mắt, chân mày, mũi, và miệng.

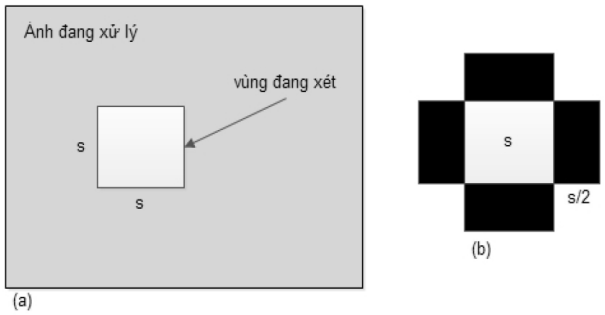


Hình 7. Rút trích vùng Ridge và Valley với mức σ đã chọn.

Trong [6], để cải thiện mức độ dò tìm thông tin *Ridge* và *Valley* trên ảnh, Hong Du đã đề xuất ý tưởng dò tìm các đặc trưng *Ridge* và *Valley* tại các mức khác nhau. Với phương pháp này, tại mức thấp thì dò tìm các đặc trưng *Ridge* và *Valley* có kích thước nhỏ và chi tiết như mắt, mũi... Trong những mức độ cao hơn kế tiếp, sẽ dò tìm các vùng *Ridge* và *Valley* lớn hơn. Đến mức lớn nhất, toàn bộ khuôn mặt có thể trở thành một vùng *Ridge*. Tức là, ở mức lớn sẽ dò được vùng *Ridge* và *Valley* kích thước lớn và ít chi tiết hơn. Tương tự như khi nhìn một người ở rất xa thì chỉ có thể thấy được dáng, gần hơn ta có thể nhìn được mặt nhưng không nhìn được các chi tiết mắt mũi, miệng, và gần hơn nữa ta mới có thể nhìn thấy rõ toàn bộ các chi tiết trên khuôn mặt.

1. Dò tìm và rút trích vùng Ridge và vùng Valley

Theo [6], để đánh giá một vùng có phải là *Ridge* hay là *Valley* hay không *Ridge* cũng không *Valley*, ta chọn bộ lọc sau (Hình 8):



Hình 8. Dò tìm vùng trên ảnh:
(a) vùng đang xét trên ảnh; (b) bộ lọc vuông.

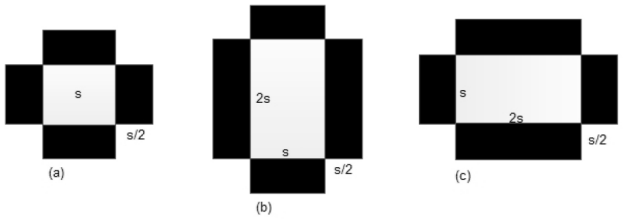
Ứng với vùng đang xét, với hình vuông s×s, tính giá trị của bộ lọc trên vùng này theo công thức (1). Đặt *S1*, *S2*, *S3*, và *S4* là các hình chữ nhật đen xung quanh hình vuông *R* (Hình 8b), giá trị lọc *F(R)* được định nghĩ như sau:

$$F(R)=2sum(R)-\sum_{i=1}^{4} sum(S_i) \qquad (1)$$

Trong công thức (2), hàm *sum(x)* tính tổng độ sáng trong vùng *x*. Dựa trên giá trị *F(R)* vừa tính được, đánh giá xem vùng *R* là *Ridge* hay *Valley* theo công thức sau:

$$R = \begin{cases} \text{Ridge} \Leftrightarrow |F(R)| \geq \delta \wedge F(R) > 0 \\ \text{Valley} \Leftrightarrow |F(R)| \geq \delta \wedge F(R) < 0 \\ \text{Normal} \Leftrightarrow (\text{vùng khác}) \end{cases} \quad (2)$$

Công thức (2) dùng để dò tìm các vùng *Ridge* và *Valley* trên ảnh. Bằng cách thay kích thước của bộ lọc, có thể dò tìm các vùng *Ridge* và *Valley* ứng với nhiều kích cở, mức độ khác nhau. Ngoài ra, để dò tìm và rút trích các vùng *Ridge* và *Valley* có hình dạng khác nhau, chúng tôi sử dụng một tập các lọc hình học có hình dạng như trong hình 9 thay vì chỉ sử dụng một bộ lọc vuông [6].



Hình 9. Tập các bộ lọc kích thước S: (a) lọc vuông; (b) lọc chữ nhật đứng; (c) lọc chữ nhật ngang.

Với tập bộ lọc được bổ sung như trên, công thức (1) được viết lại như sau:

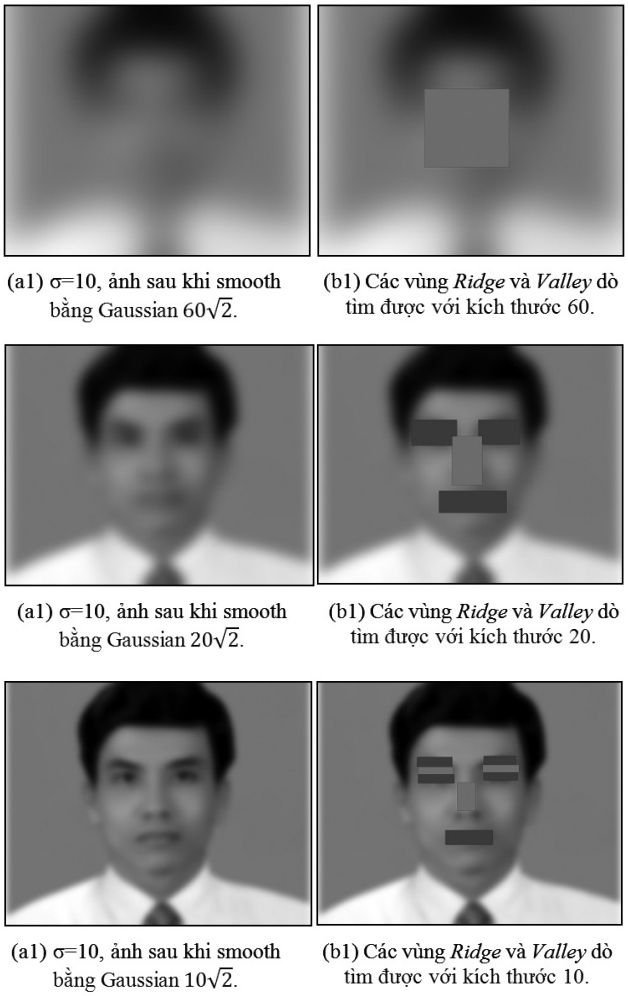
$$F_S(R) = 2\text{sum}(R) - \sum_{i=1}^{4} \text{sum}(S_i) \quad (3)$$

$$F_S(R) = 3\text{sum}(R) - 2\sum_{i=1}^{4} \text{sum}(S_i) \quad (4)$$

$F_S(R)$ được sử dụng để tính giá trị lọc vuông $F_R(R)$ sử dụng để tính giá trị các lọc hình chữ nhật. Trong công thức (3) và (4), hệ số đứng trước hàm *sum(R)* có được là do tỷ lệ giữa diện tích vùng ở tâm (vùng đang xét) và tổng diện tích của vùng lân cận. Trong Hình 9(a), tổng diện tích của các vùng đen xung quanh gấp đôi diện tích phần ở giữa, vì vậy chúng ta có hệ số 2. Trong Hình 9(b) và 9(c), tổng diện tích các vùng đen gấp rưỡi diện tích phần hình chữ nhật ở giữa. Một thông số quan trọng khác trong công thức (2) là ngưỡng δ. Ngưỡng δ phải phụ thuộc tuyến tính vào diện tích của vùng đang xét. Chọn $\delta_1 = KS^2$ đối với lọc hình vuông có kích thước S và $\delta_2 = 4KS^2$ đối với các lọc hình chữ nhật có kích thước *s*. K là một thông số thực nghiệm. Ngưỡng δ_1 là tích của *K* và s^2 là diện tích của lọc vuông.

2. Dò tìm và phát hiện vùng Ridge và Valley tại nhiều mức

Với nhiều kích thước bộ lọc khác nhau, ta sẽ dò tìm được các vùng *Ridge* và *Valley* ở nhiều mức khác nhau. Trong đó các vùng được rút mỗi mức thấp hơn sẽ được dò tìm dựa trên các vùng ở mức cao hơn.



(a1) σ=10, ảnh sau khi smooth bằng Gaussian $60\sqrt{2}$. (b1) Các vùng *Ridge* và *Valley* dò tìm được với kích thước 60.

(a1) σ=10, ảnh sau khi smooth bằng Gaussian $20\sqrt{2}$. (b1) Các vùng *Ridge* và *Valley* dò tìm được với kích thước 20.

(a1) σ=10, ảnh sau khi smooth bằng Gaussian $10\sqrt{2}$. (b1) Các vùng *Ridge* và *Valley* dò tìm được với kích thước 10.

Hình 10. Dò tìm Ridge và Valley ở nhiều mức khác nhau.

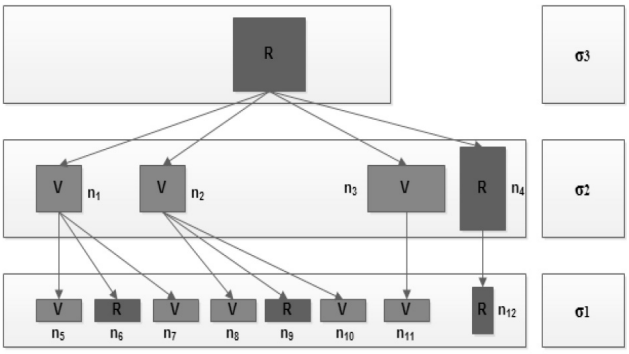
Trong Hình 10, các hình vuông và hình chữ nhật màu đỏ là vùng Ridge, còn các hình vuông và hình chữ nhật màu xanh là các vùng Valley.

Đặt $\sigma_1, \sigma_2, \ldots \sigma_n$ là các mức được cho trước và có độ lớn tăng dần từ σ_1 đến σ_n. Trong đó, σ_1 là mức có thể dò tìm các vùng *Ridge* và *Valley* có kích thước nhỏ và chi tiết nhất. σ_n là mức cao nhất mà tại đó các vùng *Ridge* và *Valley* dò tìm được chi tiết lớn và ít nhất. Tại mỗi mức, trước tiên áp dụng bộ lọc Gaussian với kích thước σ_i, sau đó sử dụng 3 bộ lọc hiệu với kích thước $s=\sigma_i$.

3. Không gian nút và khoảng cách giữa các nút

Với các vùng *Ridge* và *Valley* đã được rút trích, các vùng ở mức σ_n cao nhất, ứng với mỗi vùng như vậy ta có một "không gian nút" (*node - space*), không gian này sẽ là cơ sở để xây dựng khung lưới xác định khuôn mặt.

Với mỗi không gian nút được xây dựng, ta cấu trúc chúng thành một khung lưới để tạo thành một ứng viên cho việc xét tìm khuôn mặt trong đó. Trong không gian nút này, tâm của mỗi vùng *Ridge* và *Valley* đã rút trích được xem là một điểm (nút) trong không gian. Vùng giới hạn được xác định bởi vùng *Ridge* và *Valley* được rút trích ở mức σ_n cao nhất. Các vùng *Ridge* và *Valley* rút trích sẽ xây dựng thành không gian nút được phân chia theo mức như Hình 11.



Hình 11. Không gian nút với các nút được phân theo mức.

Trong Hình 11, vùng giới hạn được xác định là vùng *Ridge* ở mức σ_3.Với mỗi nút bất kỳ trong không gian này, chúng đều có mối liên hệ với nhau bằng khoảng cách, mức σ của nó và vùng giới hạn mà nó nằm trong đó. Để xác định được khoảng cách giữa các nút, ta xây dựng hàm khoảng cách được định nghĩa như sau:

Hàm khoảng cách: Đặt N_1 có tọa độ (x_1, y_1) và N_2 có tọa độ (x_2, y_2) là các tâm của hai nút. Khoảng cách $D(N_1, N_2)$ là hàm khoảng cách (*Distance*) giữa các nút nếu như:

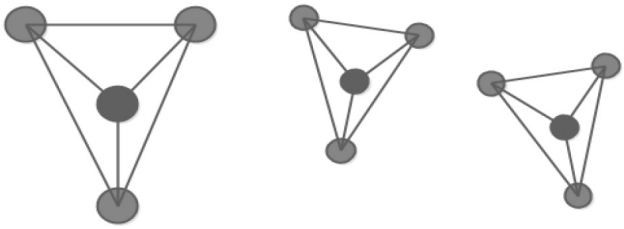
$$\begin{cases} D(N_1, N_2) \geq 0 \text{ với mọi nút thuộc không gian;} \\ D(N_1, N_2) = D(N_1, N_2) \text{ với mọi nút thuộc không gian;} \\ D(N_1, N_2) = 0 \text{ khi và chỉ khi } N_1 = N_2; \\ D(N_1, M) \leq D(N_1, N_2) + D(N_1, M); \text{ M là một nút khác} \end{cases}$$

Khoảng cách Euclide: Xét về mặt toán học, đây là khoảng cách *thường* giữa hai điểm. Do đó, khoảng cách euclide giữa hai nút $N_1 (x_1,y_1)$ và$N_2 (x_2,y_2)$ được định nghĩa như sau:

$$De(N_1,N_2)=\sqrt{(x_1- x_2)^2+(y_1- y_2)^2}$$

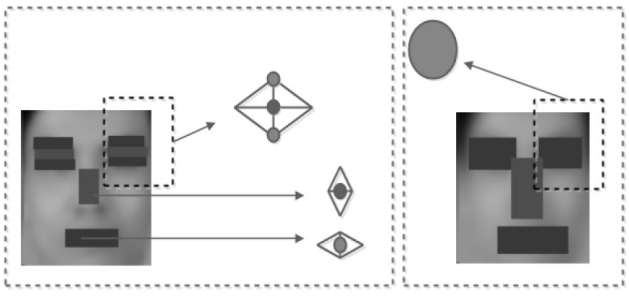
3.3. Xây dựng khung lưới mô hình hóa 3D

Tại mỗi không gian nút, dựa vào dữ liệu đã có để xây dựng một khung lưới. Trước hết, những nút ở mức σ cao sẽ được đưa vào khung lưới và nối với nhau bằng các cạnh vô hướng để tạo khung cơ bản. Kích thước của khung cơ bản này có thể lớn nhỏ tùy khoảng cách Euclide giữa các nút.



Hình 12. Khung lưới cơ bản biểu diễn không gian nút.

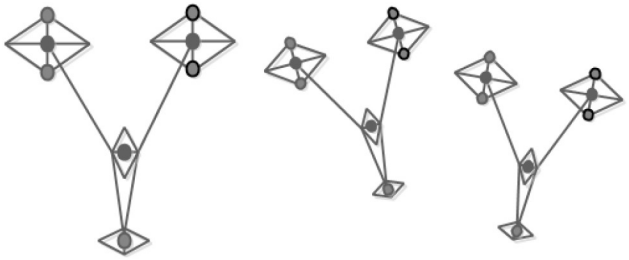
Sau đó, những nút ở mức σ thấp hơn sẽ được chèn vào trong khung và được nối với nút cha của chúng. Với các nút có cùng nút cha, nếu nút *Ridge* nằm giữa hai nút *Valley* theo chiều dọc thì ta vẽ thêm cạnh nối hai biên của nút *Ridge* này với hai nút *Valley* kia. Có thể dễ dàng nhận ra điều này sẽ tạo nên một hình thoi nằm trong vùng *Ridge* hoặc *Valley* của nút cha.



Hình 13. Các nút tạo thành từ các vùng Ridge và Valley rút trích được theo mức

Sau khi chèn hết các nút ở một mức, xóa các nút cha và các cạnh tương ứng để rút gọn khung lưới. Lúc này với mỗi vùng *Ridge* và *Valley* rút trích ở mức cao nhất sẽ đạt được một khung lưới theo định dạng 3D với độ cao vào thấp của các

nút được thể hiện qua tính chất của các nút (nút *Ridge* hay *Valley*).



Hình 14. Khung lưới chi tiết được rút trích từ các vùng Ridge và Valley.

Trong Hình 13, dễ dàng nhận thấy rằng các khung lưới tùy thuộc vào trạng thái, độ nghiêng của khuôn mặt cũng sẽ được biểu diễn với những góc nghiêng khác nhau. Điều này giúp tăng việc xác định các khuôn mặt được chụp trong ảnh hiệu quả hơn so với nhiều phương pháp khác.

4. Thực nghiệm

Với mô hình khung như trên, chúng tôi xây dựng mô hình phát hiện để dò tìm và phát hiện khuôn mặt. Bên cạnh đó, nếu có thể áp dụng thêm một số thuật toán nhằm tăng cường khả năng rút trích các vùng *Ridge* và *Valley* có độ chính xác cao hơn thì mô hình này có thể mở rộng để dùng cho bài toán nhận dạng cơ quan trên khuôn mặt chẳng hạn như mắt, mũi, chân mày...

4.1. Tập mẫu xây dựng tập khung chuẩn

Để xây dựng tập khung chuẩn, chúng tôi sử dụng tập dữ liệu ảnh khuôn mặt được tổng hợp bởi Peter Hancock [17]. Chúng tôi lựa chọn tập ảnh gồm có 131 ảnh chụp khuôn mặt của 49 người đàn ông và 20 người nữ. Các khuôn mặt được chụp ở góc nhìn trực tiếp, biểu cảm tự nhiên hoặc cười.

Việc lựa chọn tập ảnh này giúp cho quá trình trích lọc các vùng *Ridge* và *Valley* vào tạo khung lưới mô tả khuôn mặt được chính xác, không bị nhiễu.

Hình 15. Một số hình khuôn mặt trong tập ảnh của Peter Hancock.

4.2. Tập ảnh thử nghiệm

Để thử nghiệm chương trình, chúng tôi sử dụng tập ảnh chụp khuôn mặt được thu thập bởi Markus Weber [15] gồm có 450 ảnh kích thước 896×592 mm lưu ở định dạng JPEG. Đây là tập dữ liệu ảnh chụp các khuôn mặt thẳng, chụp dưới nền phức tạp, đa dạng gồm cả ngoại cảnh và trong nhà với các điều kiện chiếu sáng đa dạng...

a) Ảnh chụp trong nhà. b) Ảnh chụp ngoài trời.

Hình 16. Một số ảnh trong tập ảnh của Weber.

Bên cạnh đó chúng tôi cùng sử dụng thêm tập ảnh từ cơ sở dữ liệu ảnh của CMU/VASC [16]. Tập ảnh này gồm có 4 phần với 180 ảnh đen trắng lưu ở định dạng GIF. Các khuôn mặt được chụp trong nhiều khung cảnh khác nhau, ảnh chụp nhiều khuôn mặt, khuôn mặt chụp ở góc nghiêng hoặc ảnh chụp khuôn mặt trong tạp chí, ảnh chụp có độ nhiễu cao...

Hình 17. Một số ảnh trong tập ảnh CMU VASC.

4.3. Tạo tập khung chuẩn

Từ tập ảnh mẫu chỉ chứa khuôn mặt ở phần IV.1, chúng tôi rút trích các vùng *Ridge* và *Valley* trên ảnh. Trong thực nghiệm, chúng tôi đã lựa chọn các mức σ_1=20, σ_2=30, σ_3=120 là những mức đảm bảo độ chính xác tương đối khi rút trích các cùng *Ridge* và *Valley*.

5. Kết quả thực nghiệm

Để đánh giá kết quả thử nghiệm, từ ảnh đầu vào ta dùng phương pháp rút trích đặc trưng Ridge và Valley để nhận được một khung lưới. Dựa vào tập khung chuẩn đã xây dựng để kiểm tra xem liệu đây có phải là khung biểu diễn khuôn mặt hay không. Nếu thỏa điều kiện đánh giá nó sẽ trở khung lưới ứng viên khuôn mặt.

Bảng 1. Kết quả thử nghiệm trên các tập ảnh Weber và CMU

Tập ảnh thử nghiệm	Số ảnh	Đúng	Sai	Tỷ lệ sai	Độ chính xác
Weber	450	446	4	0.08	99,92
CMU/VASC	180	159	21	16,4	83,6

Thời gian xử lý trung bình cho mỗi ảnh ở tập Weber là 1.03s và cho tập ảnh CMU/VASC là 1.24s. Lý do cho sự khác biệt này là các ảnh trong tập Weber có độ sáng tốt hơn, độ phân giải cũng cao hơn nhiều so với các ảnh trong tập CMU/VASC, yếu tố này ảnh hưởng lớn đến khả năng xác định các vùng Ridge và Valley nhanh hay chậm và ảnh hưởng đến tính chính xác của kết quả rút trích.

Để có sự so sánh chúng tôi cũng thử kiểm tra trên một chương trình phát hiện khuôn mặt với tiếp cận dựa trên diện mạo sử dụng mạng Nơ-ron và kết quả được thể hiện ở Bảng 2.

Bảng 2. Kết quả thử nghiệm theo tiếp cận mạng Nơ-ron trên các tập ảnh Weber và CMU

Tập ảnh thử nghiệm	Số ảnh	Đúng	Sai	Tỷ lệ sai (%)	Độ chính xác (%)
Weber	450	430	20	4,5	95,5
CMU/VASC	180	135	45	25	75

Thời gian thực hiện trung bình của các lần thử nghiệm so với các tiếp cận của chúng tôi là tương đối nhiều hơn với 3.046s cho tập ảnh Weber và 4,153s cho tập CMU/VASC. Và một điều chúng tôi nhận thấy khi thử nghiệm trên chương trình phát hiện khuôn mặt sử dụng mạng Nơ-ron này là với ảnh có độ phân giải càng cao thì tốc độ xử lý của chương trình càng chậm và dẫn đến thời gian xử lý tăng lên đây là một điều tối kỵ đối với các hệ thống phát hiện và nhận dạng.

Trong một số trường hợp, việc phát hiện khuôn mặt vẫn xảy ra nhầm lẫn các vùng khác. Điều này xảy ra do ảnh thu nhận có chứa nhiều vùng ảnh có lưới *Ridge* và *Valley* có kết cấu tượng tự với kết cấu *Ridge* và *Valley* của khuôn mặt, do đó thuật toán hiểu nhầm như là một thành phần biểu diễn khuôn mặt.

6. Kết luận

Bài báo đã đề xuất và cài đặt thử nghiệm một phương pháp phát hiện khuôn mặt theo tiếp cận rút trích đặc trưng *Ridge* và *Valley*. Các kết quả thực nghiệm và so sánh cho thấy phương pháp này đạt độ chính xác cao hơn so với các hướng tiếp cận đã có.

Với khả năng mô phỏng khuôn mặt bằng một khung lưới, tiếp cận này giúp xác định được những khuôn mặt bị che khuất hoặc các hình họa khuôn mặt (không phải là khuôn mặt thật) được chụp từ người tốt hơn so với các cách tiếp cận khác.

Tài liệu tham khảo

[1] James L. Crowley, Alice C. Parker, "A Representation for Shape based on Peaks and Ridges in the Difference of Low-Pass Transform", IEEE Transaction on PAMI, pp 156-170, Vol. 6, No. 2, Mar 1984.

[2] T. K. Leung, M. C. Burl, and P. Perona, "Finding faces in cluttered scenes using random labeled graph matching," inProc. 5th International Conference on Computer Vision (ICCV '95), pp. 637–644, Boston, Mass, USA, June 1995

[3] Thanh Hai Tran Thi, Augustin Lux, "A method for Ridge Extraction", 6th Asian Conference on Computer Vision 2004 - ACCV'04, Vol. 2, Feb 2004, Korea.

[4] Paul Viola, Michael Jones, "Robust Real-time Object Detection", Vancouver, Canada, july 13, 2001.

[5] Hichem Sahbi, Donald Geman , "*A Hierarchy of Support Vector Machines for Pattern Detection*", Journal of Machine Learning Research 7 (2006) 2087-2123.

[6] Tran Le Hong Du, Duong Anh Duc, Duong Nguyen Vu, "*Ridge and Valley based Face Detection*", IEEE International Conference on Computer Sciences Research, Innovation, Vision for the Future (RIVF'06), Ho-Chi-Minh city, Vietnam, 2006.

[7] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation,"IEEE Trans. Pattern Anal. Machine Intell., vol. 19, no. 7, pp. 696–710, 1997.

[8] E. Osuna, R. Freund, F. Girosi, "Training Support Vector Machines: An Application to Face Detection", International Conference on Computer Vision and Pattern Recognition -CVPR'97, pp. 130-136, 1997.

[9] H. Rowley, S. Baluja, T. Kanade, "Neural network-based face detection", IEEE Transaction on Pattern Analysis and Machine Intelligence (PAMI), Vol. 20, No. 1, pp.23-38, Jan 1998.

[10] F. Samaria, Face recognition using hidden Markov models,Ph.D. dissertation, University of Cambridge, Cambridge, UK,1994.

[11] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: a literature survey," ACM Computing Surveys, vol. 35, no. 4, pp. 399–458, 2003.

[12] C. Canton-Ferrer, J. R. Casas, and M. Pardàs, "Fusion of multiple viewpoint information towards 3D face robust orientation detection", IEEE International Conference on Image Processing, 2005, Vol. 2, pp. 366–369.

[13] Quan Yuan, Wen Gao, Hongxun Yao, "Robust frontal face detection in complex environment", International Conference on Pattern Recognition 2002 – ICPR'02, pp 25-28, Aug 2002, Canada.

[14] Phạm Thế Bảo, Nguyễn Thành Nhựt, Cao Minh Thịnh, Trần Anh Tuấn, Phan Phúc Doãn, "Tổng quan các phương pháp xác định khuôn mặt người", Tạp chí Công nghệ thông tin & Truyền thông, 2007.

[15] Markus Weber, "Frontal face dataset", California Institute of Technology, 2003.http://www.vision.caltech.edu/html-files/archive.html.

[16] CMU/VASC Image Database, Vision and Autonomous Systems Center, Carnegie Mellon University, 2003.

[17] Peter Hancock, Psychological Image Collection at Stirling (PICS), Psychology, School of Natural SciencesUniversity of Stirling, 2008.http://pics.psych.stir.ac.uk/2D_face_sets.htm.

[18] Luxand FaceSDK 4.0, Trial Version, Luxand. Inc,2011.http://www.luxand.com.