

Nhận dạng khuôn mặt trực tiếp và ứng dụng trong mở cửa tự động

Live Face Recognition and Its Application in Automatic Door Locking System

Phạm Văn Trường^{1*}, Trần Thị Thảo²

ĐH Bách Khoa Hà Nội

Email: truong.phamvan@hust.edu.vn

Abstract

This paper proposes an approach for live face recognition on Raspberry Pi, and used for face recognition based door locking system. The face images are acquired by an camera and fed to live face detection and face recognition system. The face detection is based on Haar cascades to extract face features. Then the face is analyzed to check whether it is a live face or a fake face image fed from a printed picture or video containing face image sequences. In this study, we consider the information from the changes of pixel intensity in a series of acquired face images that reflect the changes of heart rate during the rhythmic contraction and relaxation of the heart. Once verified as a live face, the face image is fed to the face recognition process using eigenfaces and principal component analysis. The input image is compared with the database to check whether it belongs to a person in the database or not. If yes, the door is controlled to allow an entrance. The live face detection and recognition was validated on a database including 25 people with 400 training images. The average accuracy obtained approximately 80%. The hardware including camera and actuators are with low expense but meet real time performance.

Keywords: Live face detection, face recognition, principal component analysis (PCA), PPG, Raspberry Pi, door locking system

Symbols

Symbols	Units	Description
C		Covariance matrix
Y		Input vector
T		Linear transformation

Abbreviations

PCA	Principal Component Analysis
MSE	Mean Squared Error
ICA	Independent Component Analysis
PPG	Photoplethysmography
OpenVC	Open Source Computer Vision

Tóm tắt

Bài báo này đề xuất một phương pháp phát hiện khuôn mặt trực tiếp và nhận dạng người thực hiện trên máy tính nhúng Raspberry Pi. Các ảnh đầu vào được thu thập từ camera và đưa vào thuật toán phát hiện khuôn mặt trực tiếp và nhận dạng khuôn mặt. Phân phát hiện khuôn mặt sử dụng Haar cascades để trích xuất ra các đặc trưng của khuôn mặt. Tiếp đến, để kiểm tra xem khuôn mặt đưa vào camera có thật sự là khuôn mặt người thực tế đang chụp hay là khuôn mặt giả

từ ảnh in hoặc video đưa vào, một phương pháp xác định khuôn mặt trực tiếp dựa trên sự thay đổi cường độ sáng trên khuôn mặt nhờ thay đổi nhịp tim ở các thời điểm khác nhau trên khung hình. Sau khi đã xác định được đúng khuôn mặt thật, hình ảnh được đưa vào nhận diện thông qua thuật toán phân tích thành phần chính dựa trên khuôn mặt riêng eigenfaces nhờ so sánh với cơ sở dữ liệu đã có. Nếu đúng khuôn mặt nằm trong dữ liệu thuộc đối tượng cho phép truy nhập, bộ điều khiển sẽ thực hiện thao tác mở cửa. Thuật toán đã được thực hiện trên cơ sở dữ liệu gồm 25 người với 400 ảnh chụp huấn luyện, cho độ chính xác 80%, và thực thi trên máy tính với camera và cơ cấu chấp hành có giá thành hạ nhưng vẫn đáp ứng thời gian thực.

Chữ viết tắt

PCA	Principal Component Analysis
MSE	Mean Squared Error
ICA	Independent Component Analysis
PPG	Photoplethysmography
OpenCV	Open Source Computer Vision

1. Phần mở đầu

Trên đà phát triển của cuộc cách mạng công nghiệp 4.0, trí tuệ nhân tạo nói chung và thị giác máy tính nói riêng hiện

đã và đang là xu hướng nghiên cứu được phát triển và ứng dụng ngày càng rộng rãi trên toàn thế giới. Thị giác máy tính dần được ứng dụng vào các lĩnh vực như robot, tự động hóa, bảo mật và y sinh... mang lại hiệu quả với độ chính xác cao. Các ứng dụng trong bảo mật sử dụng hình ảnh có thể kể tới như sinh trắc học vân tay, móng mắt [1], lòng bàn tay [2], và nhất là nhận dạng khuôn mặt [3, 4] ngày càng chiếm lĩnh vị trí quan trọng trong mọi mặt của đời sống, kinh tế cũng như an ninh, quốc phòng... Trong đó, nhận diện khuôn mặt được ứng dụng rộng rãi nhờ những ưu điểm nổi bật như dễ dàng thu thập hình ảnh không tiếp xúc, dễ nhận biết và kiểm chứng một cách trực quan.

Nhận diện khuôn mặt là công nghệ sinh trắc học ít xâm nhập và nhanh nhất, hoạt động với định danh cá nhân rõ ràng nhất - khuôn mặt con người. Thay vì yêu cầu mọi người tương tác lên thiết bị như đặt tay lên máy quét hoặc định vị chính xác mắt của họ trước máy quét, hệ thống nhận diện khuôn mặt kín đáo chụp ảnh khuôn mặt của mọi người nhập vào một khu vực xác định và trong hầu hết các trường hợp, các đối tượng hoàn toàn không biết về quy trình chụp ảnh và nhận diện này. Do vậy họ không cảm thấy bị giám sát hoặc sự riêng tư của họ đã bị xâm phạm.

Hệ thống nhận diện khuôn mặt được ứng dụng trong rất nhiều lĩnh vực từ an ninh quốc phòng đến đời sống kinh tế. Đặc biệt trong lĩnh vực an ninh công cộng, hệ thống này được sử dụng để: Giám sát bằng camera để phát hiện tội phạm tại các khu vực công cộng, tra cứu thông tin tội phạm, tìm trẻ lạc dựa vào các camera đặt ở các nơi công cộng. Trong kinh tế, đời sống, nhận diện khuôn mặt ngày nay càng được ứng dụng nhiều trong các ứng dụng an ninh và thông tin giải trí như hệ thống phát hiện các nhân vật VIP đặt chân vào khách sạn, xác định người nổi tiếng trong các video hay bức ảnh, đặt trong shop để nhận diện khách quen hay theo dõi kẻ gian... Ngày nay hệ thống nhận diện khuôn mặt được ứng dụng rộng rãi trong các ứng dụng dân sự như hệ thống điểm danh trong công ty, mở cửa ở các tòa nhà thông minh...

Đã có nhiều công trình nghiên cứu về bài toán xác định khuôn mặt và nhận diện khuôn người từ ảnh đen trắng, ảnh xám, đến ảnh màu [3, 5]. Các nghiên cứu đi từ bài toán đơn giản là ảnh chỉ chứa một khuôn mặt người nhìn thẳng vào thiết bị thu hình và đầu ở tư thế thẳng đứng, cho đến ảnh màu với nhiều khuôn mặt người trong cùng ảnh, khuôn mặt có quay một góc nhỏ, hay bị che khuất một phần, và với ảnh nền của ảnh phức tạp nhằm đáp ứng nhu cầu thật sự cần thiết của con người. Bài toán nhận diện khuôn mặt thường gồm hai bước chính: xác định khuôn mặt và nhận dạng mặt. Bước xác định khuôn mặt người là một kỹ thuật máy tính để xác định các vị trí và các kích thước của khuôn mặt trong ảnh bất kỳ. Kỹ thuật này nhận biết các đặc trưng khuôn mặt và bỏ qua những thứ khác như: toà nhà, cây cối, cơ thể... Còn bước nhận dạng mặt người là xác định danh tính tự động cho từng ảnh đối tượng người dựa vào nội dung của ảnh và cơ sở dữ liệu của khuôn mặt người đã có.

Hiện nay đã có nhiều thuật toán khác nhau để xác định khuôn mặt cũng như nhận diện khuôn mặt, mỗi phương pháp có ưu nhược điểm riêng. Một vài phương pháp sử dụng đường viền, một số khác thậm chí phức tạp hơn như sử dụng mạng neural, bộ lọc... [6, 7]. Những thuật toán đối mặt với những vấn đề khác nhau, nhất là sự hài hòa giữa độ chính xác và độ phức tạp tính toán. Thực tế cho thấy, thuật toán cho độ chính xác càng cao thì thường độ tính toán càng phức

tạp và yêu cầu cao về đáp ứng phần cứng. Vì vậy một vấn đề đặt ra trong thực tế, nhất là trong các ứng dụng nhỏ yêu cầu giá thành thấp, là làm thế nào vừa lựa chọn được thuật toán xử lý nhanh, thực hiện trên các máy tính cấu hình vừa phải mà vẫn đạt được độ chính xác chấp nhận được. Nghiên cứu này tập trung vào bài toán thực tế là đưa ra giải pháp cho một bài toán nhận diện khuôn mặt đáp ứng thời gian thực trên máy tính nhúng có giá thành thấp. Với bài toán xác định khuôn mặt, các tác giả sử dụng thuật toán Haar Cascades để xác định khuôn mặt và thuật toán Eigenfaces để nhận diện khuôn mặt. Việc xác định khuôn mặt bằng Haar Cascades cho hiệu quả cao, việc xác định được nhiều khuôn mặt trên cùng một hình ảnh, tránh bỏ sót khuôn mặt cần xử lý.

Mặt khác, hiện nay trong hầu hết các nghiên cứu về nhận diện khuôn mặt, các đề xuất tập trung chủ yếu vào bài toán tăng độ chính xác của việc nhận diện, định danh chính xác một người nào đó trong cơ sở dữ liệu, mà ít đề cập đến vấn đề nhận diện trực tiếp. Thực tế, việc phát hiện khuôn mặt trực tiếp có ý nghĩa khi cần chứng thực hình ảnh tương ứng là của người đang xét, cần định danh chứ không phải sao chép của người đó, ví dụ do ảnh in ra mình được đưa vào máy quét. Việc phát hiện trực tiếp này có ý nghĩa quan trọng trong việc loại trừ được khả năng có kẻ xấu lợi dụng có được hình ảnh của chủ thể của thành viên để truy nhập vào hệ thống một cách trái phép. Các nghiên cứu về phát hiện trực tiếp khuôn mặt trước khi đưa vào nhận diện trong các ứng dụng định danh còn ít được quan tâm. Phương pháp chủ yếu được sử dụng hiện nay là nhận biết thông qua sự chớp mắt [8, 9]. Phương pháp này tuy đơn giản nhưng thường có nhược điểm là cần thu thập hình ảnh trong một thời gian đủ để ước lượng được khả năng có sự chớp mắt xảy ra, dẫn đến khả năng chậm trễ trong việc xử lý của hệ thống. Trong nghiên cứu này, chúng tôi đề xuất một cách tiếp cận khác, dựa trên ý tưởng về đo nhịp tim không tiếp xúc thông qua sự thay đổi cường độ sáng của ảnh. Cụ thể, chúng tôi chỉ kiểm tra sự thay đổi cường độ sáng theo sự thay đổi lưu lượng máu trong một thời gian ngắn, chỉ vài giây mà vẫn phát hiện được người trực tiếp.

Với bài toán nhận dạng khuôn mặt của mình, yêu cầu để đáp ứng thời gian thực, chúng tôi lựa chọn thuật toán Eigenfaces, đây là phương pháp cơ bản, dễ dàng nhận diện được khuôn mặt ở một khoảng cách gần, cường độ ánh sáng đủ lớn. Thuật toán chạy trên hệ máy tính nhúng nhỏ gọn Raspberry Pi. Thông qua camera, hình ảnh xác định có khuôn mặt sẽ được hệ thống nhúng xử lý, nhận diện, định danh từ đó gửi đi các lệnh để điều khiển cơ cấu chấp hành chính là một động cơ. Ứng dụng điều khiển động cơ thông qua một Board điều khiển nhận lệnh từ hệ thống nhúng. Ngoài ra còn có khối nguồn cung cấp nguồn cho hệ thống nhúng, Board điều khiển và động cơ Servo.

2. Xác định khuôn mặt trực tiếp và nhận dạng

2.1. Xác định khuôn mặt

Trong nghiên cứu này, bước đầu tiên sau khi thu thập được ảnh từ camera là xác định được khuôn mặt trong bức ảnh. Mục đích của bước này là tách ra khỏi khung nền, phục vụ cho việc xử lý của khối nhận dạng phía sau, chính vì điều này ta có thể sử dụng bộ phân loại Haar cho việc xác định

khuôn mặt, điều này cho phép xác định nhanh và chính xác vùng khuôn mặt, phục vụ tốt cho việc nhận dạng.

Hệ thống xác định đối tượng là cho một mảng hình ảnh đã biết kích thước để quyết định xem các đặc tính này xuất phát từ một đối tượng không. Với mục đích cho việc thực hiện xác định đối tượng, Haar Classifier [3, 5] là một lựa chọn hợp lý được áp dụng cho hệ thống này. Haar Classifier mã hóa sự tồn tại của sự tương phản định hướng giữa các vùng trong hình ảnh. Một thiết lập của những đặc tính có thể được sử dụng để mã hóa độ tương phản được thể hiện bởi một đối tượng.

Đặc trưng Haar được dùng để phân loại hình ảnh đối tượng. Tập đặc trưng này xem xét những vùng hình chữ nhật của hình ảnh và tổng hợp những điểm ảnh trong vùng này. Giá trị thu được sử dụng để phân loại hình ảnh. Ví dụ, ta có một cơ sở dữ liệu hình ảnh với khuôn mặt người và các tòa nhà. Điều đó có thể nếu vùng mắt và vùng tóc của khuôn mặt được xem xét, thì tổng các điểm ảnh trong vùng này sẽ khá cao đối với khuôn mặt người và tùy ý cao hoặc thấp cho các tòa nhà. Giá trị thứ hai sẽ phụ thuộc vào cấu trúc của tòa nhà, môi trường của nó, trong khi các giá trị của nó sẽ ít nhiều giống nhau. Do đó, chúng ta có thể phân loại tất cả các hình ảnh có đặc trưng Haar (Haar-features) giống nhau trong vùng hình chữ nhật này trong một phạm vi giá trị nhất định như là một loại và những vùng nằm ngoài phạm vi này ở vùng khác. Điều này gần như có thể phân chia bộ hình ảnh thành những ảnh có rất nhiều khuôn mặt, một vài tòa nhà hay các tòa nhà khác. Việc này có thể thực hiện lặp đi lặp lại để phân chia các nhóm hình ảnh.

Các đặc tính Haar được sử dụng để phát hiện sự hiện diện của tính năng đó trong hình ảnh nhất định. Mỗi đặc tính kết quả trong một giá trị đơn thu được bằng cách trừ tổng các điểm ảnh dưới hình chữ nhật màu trắng từ tổng các điểm ảnh dưới hình chữ nhật màu đen. Áp dụng các đặc tính Haar vào hình ảnh (đặc tính biên, đặc tính đường, đặc tính trung tâm) như trong hình 1.

Viola và Jones [3] đã sử dụng một cửa sổ 24x24 như một kích thước cửa sổ cơ bản để bắt đầu ước lượng những đặc tính này trong một hình ảnh bất kỳ đã cho. Nếu chúng ta xem xét tất cả những tham số có thể có của các đặc tính Haar như quy mô vị trí và tính toán.

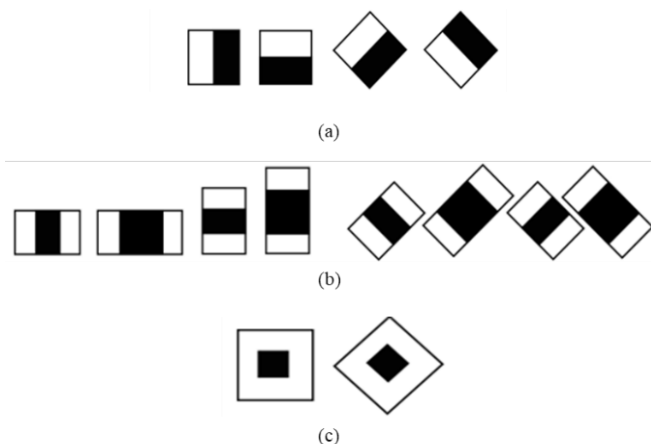


Figure 1: Các đặc tính chính của phân loại Harr. (a) đặc tính biên; (b) đặc tính đường; (c) đặc tính trung tâm.

có khoảng 160 nghìn đặc tính trong cửa sổ. Mô hình phân loại tầng được sử dụng bao gồm các giai đoạn chứa một bộ

phân loại mạnh. Vì vậy, tất cả các tính năng được nhóm thành nhiều giai đoạn, trong đó mỗi giai đoạn có một số tính năng nhất định. Công việc của từng giai đoạn được sử dụng để xác định liệu một cửa sổ phụ có chắc chắn không phải là khuôn mặt hay không. Một cửa sổ phụ đã được loại bỏ ngay lập tức nếu nó không chọn lọc thành công trong bất kỳ giai đoạn nào.

2.2. Xác định người trực tiếp thông qua nhịp tim

Sau khi đã tách được khuôn mặt từ các bức ảnh thu thập được như ở bước xác định khuôn mặt nói trên, ta cần kiểm tra xem khuôn mặt đưa vào camera có thật sự là khuôn mặt người thực tế đang chụp hay không. Điều này thực hiện nhằm tránh tình huống có sự gian lận khi có đối tượng cố tình đưa vào camera một hình ảnh khuôn mặt của người trong cơ sở dữ liệu nhưng thực ra là ảnh in hoặc video chứ không phải là một người thực tế cần chụp. Để tránh tình huống sử dụng ảnh giả, một số hệ thống yêu cầu người chụp thực hiện biểu cảm như mỉm cười, chớp mắt, ngoảnh đầu qua lại. Một số hệ thống khác thực hiện thêm một bước nhận diện cảm xúc, do vậy có thể làm cho hệ thống trở nên phức tạp hơn. Một trong số các phương pháp được sử dụng nhiều là phương pháp phát hiện chớp mắt [9]. Phương pháp này có ưu điểm là nhanh do thuật toán đơn giản nhưng sẽ khó có hiệu quả trong trường hợp không phát hiện chớp mắt được do phản xạ do kính, ánh sáng hoặc nếu người thực không thực hiện việc chớp mắt trong một thời gian vài giây.

Trong nghiên cứu này, để phát hiện người trực tiếp, các tác giả đề xuất tính toán ra nhịp tim từ sự thay đổi cường độ sáng trên khuôn mặt ở các thời điểm liên tiếp thông qua các ảnh chụp. Ý tưởng này xuất phát từ quá trình tìm hiểu các nghiên cứu gần đây cho thấy kết quả khả quan của phương pháp ước lượng nhịp tim một cách gián tiếp thông qua sự thay đổi cường độ sáng trên khuôn mặt trong một chu kỳ tuần hoàn của tim [10, 11]. Theo các nghiên cứu này, các giá trị biểu diễn sự thay đổi của cường độ sáng trên khuôn mặt ở các thời điểm khác nhau có sự tương quan với tín hiệu đo xung nhịp tim từ thiết bị đo quang tử PPG (Photoplethysmography). Đây là phương pháp đo đơn giản không xâm lấn còn gọi là nồng độ máu thông qua những biến động giữa truyền và phản xạ ánh sáng. Kỹ thuật quang điện có thể cung cấp những thông tin giá trị về hệ thống tim và mạch máu như là nhịp tim, sự bão hòa oxy trong máu, huyết áp, tín hiệu nhịp tim. Đặc biệt, PPG luôn luôn sử dụng nguồn ánh sáng (ví dụ như đỏ và/ hoặc dưới đỏ) nhưng những nghiên cứu gần đây đã cho thấy rằng kỹ thuật đo nhịp tim có thể đạt được sử dụng máy ảnh số với ánh sáng thông thường xung quanh như là nguồn ánh sáng. Tuy nhiên, PPG được biết đến là một phương pháp rất nhạy cảm với nhiễu chuyển động và việc khắc phục những vấn đề về nhiễu này là một trong những vấn đề thách thức nhất. Một kỹ thuật để loại bỏ nhiễu từ tín hiệu sinh lý đó là kỹ thuật tách nguồn mù bằng phân tích thành phần độc lập (ICA).

Trong bài toán xác định nhịp tim sử dụng PPG và theo sự thay đổi cường độ sáng trong khuôn mặt thì việc xử lý tín hiệu là một bước quan trọng để xác định chính xác nhịp tim của đối tượng cần nghiên cứu. Tuy nhiên trong nghiên cứu này, chúng ta chỉ sử dụng nguyên lý của PPG để xác định người thực thông qua việc kiểm tra sự thay đổi của cường độ sáng trong một vài giây. Vì vậy ta sẽ không cần các phương pháp hậu xử lý tín hiệu phức tạp. Thay vào đó ta chỉ cần

kiểm tra xem nếu số lần thay đổi cường độ sáng trong một giây vào khoảng 5 đến 10 lần thì ta có thể kết luận đó là người thực. Việc giám sát sự thay đổi cường độ sáng này được minh họa trong hình 2. Trong ví dụ này, ta giả sử thu thập được các hình ảnh khuôn mặt từ video dài 6 giây, với tổng cộng 180 frame. Trên hình 2 minh họa một số hình ảnh khuôn mặt ở giây thứ ba. Ở mỗi khuôn mặt, hình ảnh màu RGB (Red-Green-Blue) sẽ được tách ra thành ba kênh, Red, Green, và Blue.

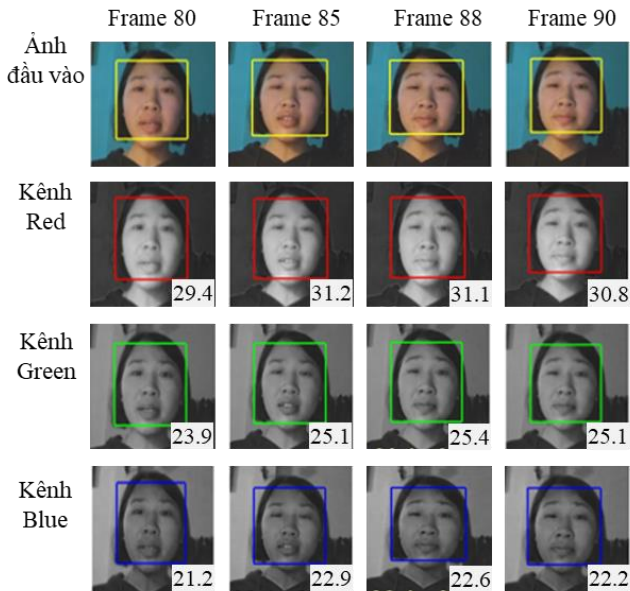


Figure 2: Minh họa sự thay đổi mức xám trung bình trong khung hình khuôn mặt theo thời gian. Video minh họa trong 6 giây gồm 180 frame. Các giá trị mức xám tương ứng trên các ảnh từng kênh được thể hiện trên mỗi hình tương ứng.

Sau đó giá trị trung bình của mức xám trên khuôn mặt (trong khung chữ nhật) ở từng kênh sẽ được tính toán. Như ta đã thấy trên hàng thứ 2, 3, và 4, cường độ sáng thông qua mức xám của khuôn mặt trong một giây đã có sự thay đổi dù mắt thường của chúng ta khó nhận biết được.

Sự thay đổi cường độ sáng này là do có sự thay đổi khi tim co bóp, chính vì vậy đồ thị mức xám của ảnh chính là đồ thị của nhịp tim. Sự thay đổi của cường độ xám ở trên các kênh trong ví dụ này được biểu diễn trên hình 3. Qua hình này ta thấy các giá trị mức xám đã thay đổi theo thời gian tương ứng với chu kỳ tuần hoàn của tim. Trong thực thi, ta có thể xác định người trực tiếp thông qua số lần thay đổi cường độ sáng của các frame bằng cách lấy trung bình số lần thay đổi của cả ba kênh Red, Green, và Blue.

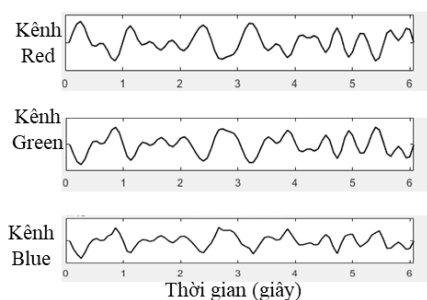


Figure 3: Các tín hiệu thể hiện sự thay đổi mức xám trung bình trong khung hình khuôn mặt theo thời gian trên từng kênh ở hình 2.

2.3. Nhận dạng khuôn mặt trong cơ sở dữ liệu

Sau khi xác định được khuôn mặt là khuôn mặt thật, ta sẽ chuyển sang công đoạn nhận diện, định danh khuôn mặt đó dựa trên phương pháp nhận diện tổng thể mà cụ thể là thuật toán sử dụng Eigenfaces - khuôn mặt riêng.

Kỹ thuật Eigenfaces (khuôn mặt riêng) sử dụng việc phân tích những thành phần chính của hình ảnh trên khuôn mặt. Việc phân tích này giảm thiểu được số kích thước của tập dữ liệu giúp thuận tiện hơn cho việc nhận diện. Eigenfaces là một tập của những vector riêng được sử dụng trong bài toán thị giác máy tính trong nhận diện khuôn mặt người. Hướng giải quyết sử dụng eigenfaces cho nhận dạng đã được phát triển bởi Sirovich và Kirby [12] và sau đó được Matthew Turk và Alex Pentland sử dụng trong phân loại khuôn mặt [13] được coi như là sự thành công đầu tiên trong công nghệ nhận diện khuôn mặt. Những vector riêng được tính toán từ ma trận hiệp phương sai phân phối xác suất của không gian vector kích thước nhiều chiều của những khuôn mặt người có thể chấp nhận. Việc coi một bức ảnh đầu vào là một vector đa chiều sẽ giúp cho việc xử lý bức ảnh trở thành việc xử lý tính toán trên vector trong đại số tuyến tính. Do ảnh được biểu diễn trong không gian hai chiều thường có kích thước lớn, để cho việc tính toán có hiệu quả, cần thiết phải có bước giảm số chiều vector. Thực tế thường sử dụng phương pháp thành phần chính PCA (Principal Component Analysis).

Mục tiêu của phương pháp PCA là giảm số chiều của một tập các vector sao cho vẫn đảm bảo được tối đa các thông tin quan trọng nhất. Điều này có nghĩa là vẫn giữ được K đặc tính mới từ các thông tin chứ không phải giữ nguyên K đặc tính ban đầu, những đặc tính mới này được suy ra từ những đặc tính cũ. Nói một cách ngắn gọn, mục tiêu của phương pháp PCA là tìm ra một không gian với số chiều nhỏ hơn không gian cũ mà các trục tọa độ của không gian mới được xây dựng sao cho trên mỗi trục, độ biến thiên dữ liệu là lớn nhất có thể. Một điểm nữa là các trục tọa độ trong không gian mới luôn đảm bảo trực giao đôi một với nhau.

Vector \mathbf{x} ban đầu có N chiều, vector \mathbf{y} mới được tạo thành chỉ còn K chiều. Khi loại bỏ một số thành phần của vector \mathbf{x} để thu được vector \mathbf{y} sẽ gây ra sai số, phương pháp PCA sẽ cố gắng tìm phép biến đổi tuyến tính \mathbf{T} thỏa mãn $\mathbf{y} = \mathbf{T}\mathbf{x}$ sao cho trung bình bình phương lỗi (MSE) là bé nhất.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (1)$$

Ở đây $\hat{\mathbf{Y}}$ là một vector với n chiều (dự đoán), \mathbf{Y} là vector đầu vào ban đầu.

Để tìm được phép biến đổi \mathbf{T} , trước hết ta gọi \mathbf{M} là vector trung bình của tất cả các vector \mathbf{x} trong tập mẫu. Ma trận hiệp phương sai của các phần tử x trong tập mẫu là \mathbf{C} , được tính theo công thức:

$$\mathbf{C} = \frac{1}{M} \sum_{k=1}^M (\mathbf{x}_k - \mathbf{M})(\mathbf{x}_k - \mathbf{M})^T \quad (2)$$

Người ta đã chứng minh được rằng: Nếu \mathbf{T} là ma trận m hàng, mỗi hàng là một vector riêng của \mathbf{C} , đồng thời m vector riêng này phải ứng với m trị riêng lớn nhất, khi đó \mathbf{T} chính là phép biến đổi thỏa mãn sai số MSE nhỏ nhất.

Phương pháp PCA quy về việc tìm giá trị riêng và vector riêng của ma trận hiệp phương sai \mathbf{C} của tập mẫu \mathbf{X} . Sau đó

ta chỉ giữ lại K vector riêng lớn nhất để làm cơ sở cho không gian mới. Các bước thực hiện của PCA có thể được liệt kê như sau:

Giả thiết x_1, x_2, \dots, x_M là M vector N chiều.

Bước 1: Biểu diễn mọi ảnh I_i thành vector Γ_i

Bước 2: Tính vector khuôn mặt trung bình (mean face) Ψ theo công thức:

$$\Psi_i = \frac{1}{M} \sum_{i=1}^M \Gamma_i \quad (3)$$

Bước 3: Trừ vector khuôn mặt trung bình (mean face):

$$\Phi_i = \Gamma_i - \Psi \quad (4)$$

Bước 4: Tính ma trận hiệp phương sai:

$$C = \frac{1}{M} \sum_{n=1}^M \Phi_n \Phi_n^T = \frac{1}{M} A A^T \quad (5)$$

Trong đó $A = [\Phi_1 \Phi_2 \dots \Phi_M]$, C là ma trận cấp $N^2 \times N^2$

Sau khi đã tính được các vector v_i (có kích thước $M \times 1$), ta sẽ dễ dàng suy ra được các vector riêng u_i (kích thước $N^2 \times 1$), mong muốn cần tìm theo công thức:

$$u_i = A v_i \quad (6)$$

Bước 5: Chỉ giữ lại K vector riêng trong số M vector nói trên (ứng với K trị riêng lớn nhất), tất nhiên $K \ll N^2$.

Có thể chọn K theo công thức sau:

$$\frac{\sum_{i=1}^K \lambda_i}{\sum_{i=1}^N \lambda_i} > \text{Ngưỡng chọn (thường từ 0.9 đến 0.95)}.$$

Sau khi đã có các vector riêng nhờ PCA, ta có thể biểu diễn các khuôn mặt có sẵn vào không gian vector như sau: Mỗi khuôn mặt Φ_i trong tập huấn luyện có thể được biểu diễn lại thành tổ hợp tuyến tính của K vector riêng có trị riêng lớn nhất

$$\Phi_i = \sum_{j=1}^K w_j u_j \quad (7)$$

Trong đó: $w_j = u_j^T \Phi_i \quad (8)$

Ta gọi các vector riêng u_i là các Eigenfaces (khuôn mặt riêng). Như vậy, từ bây giờ, mỗi khuôn mặt trong tập huấn luyện Φ_i sẽ được biểu diễn trong không gian mới:

$$\Omega_i = \begin{bmatrix} w_1^i \\ w_2^i \\ \dots \\ w_K^i \end{bmatrix} \quad \text{Với } i=1, 2, \dots, M$$

$$\text{Cụ thể ta có: } \Omega_i = \begin{bmatrix} u_1^T \cdot \Phi_i \\ u_2^T \cdot \Phi_i \\ \dots \\ u_K^T \cdot \Phi_i \end{bmatrix}$$

Để nhận diện một khuôn mặt mới, ta gọi đó là khuôn mặt chưa xác định được danh tính Γ (ảnh chính diện được đưa về kích thước giống trong tập mẫu). Giai đoạn nhận dạng

này giống hết giai đoạn biểu diễn ảnh khuôn mặt trong tập mẫu. Ta thực hiện lần lượt các bước sau:

Bước 1: Chuẩn hóa Γ : $\Phi = \Gamma - \Psi \quad (9)$

Điều này cũng tương ứng như quy vector đầu vào về không gian với gốc tọa độ mới (gốc tọa độ này gọi là meanface).

Bước 2: Biểu diễn Φ thành Ω như sau:

$$\Omega = \begin{bmatrix} u_1^T \cdot \Phi \\ u_2^T \cdot \Phi \\ \dots \\ u_K^T \cdot \Phi \end{bmatrix}$$

Bước 3:

$$\text{Tìm } e_l = \min_l \|\Omega - \Omega^l\| \quad (10)$$

Tức là ta tìm khuôn mặt thứ l trong tập mẫu có khoảng cách gần nhất với khuôn mặt cần nhận diện.

Bước 4: Nếu $e_l < T_r$ (T_r là một ngưỡng chọn chấp nhận được nào đó).

Tức là ảnh khuôn mặt cần xác định “đủ gần” với ảnh của người thứ l trong tập mẫu. Khi đó, ta kết luận đó chính là khuôn mặt của người thứ l .

3. Kết quả thử nghiệm

3.1. Thiết bị phần cứng

Các thiết bị phần cứng được thiết kế với mục đích ứng dụng vào đóng mở cửa tự động nhờ nhận diện khuôn mặt với giá thành thấp, nhỏ gọn có thể áp dụng trong thực tế. Ứng dụng có tích hợp vi điều khiển để có thể thay đổi các thuật toán linh hoạt và tiện ích phù hợp hơn với nhu cầu và công nghệ hiện đại thời nay.

Ứng dụng nhận diện khuôn mặt trong đóng mở cửa tự động hoạt động dựa trên nguyên lý là sẽ định danh khuôn mặt xuất hiện trong khung hình camera từ đó cho phép mở rồi đóng cửa tự động nếu khuôn mặt xuất hiện trong khung hình camera là thành viên có trong cơ sở dữ liệu của ứng dụng. Đồng thời thiết kế một giao diện hiển thị trên màn hình kết nối với bộ xử lý trung tâm cho người dùng có thể điều khiển và tự thêm thành viên mới vào trong cơ sở dữ liệu. Đi kèm ứng dụng là một số chức năng và tiện ích được thiết kế nhằm mục đích tiện lợi, phù hợp với việc sử dụng như nhận thông báo qua tin nhắn và server khi có thành viên trong cơ sở dữ liệu truy nhập mở cửa, quản lý dữ liệu hay có thể đóng mở cửa tự động từ xa qua tin nhắn.

Các khối chính của hệ thống đóng mở cửa dựa trên nhận diện khuôn mặt được mô tả trên hình 4. Các thiết bị chính gồm có: Camera, máy tính nhúng, màn hình giao tiếp với admin/user, boad điều khiển, và khối cơ cấu chấp hành. Camera có chức năng thu thập hình ảnh, khung hình gửi về máy tính nhúng để xử lý, ở đây ta có thể sử dụng Raspberry Pi camera module. Bộ điều khiển trung tâm ở đây sử dụng máy tính nhúng Raspberry Pi. Khối hệ thống nhúng gồm bo mạch Raspberry Pi và màn hình hiển thị đảm nhận việc thực thi thuật toán xử lý ảnh bao gồm xác định khuôn mặt và nhận diện khuôn mặt thông qua thư viện OpenCV. Hệ thống nhúng cũng đảm nhiệm việc gửi đi các lệnh làm việc với chuẩn giao tiếp UART thông qua các khối vào ra được tích hợp thiết kế có sẵn với Board điều khiển để thực thi điều khiển cơ cấu chấp hành như động cơ servo, còi chirp....

Board điều khiển sử dụng chip STM32F103C8T6 đảm nhiệm công việc xử lý lệnh nhận được từ máy tính nhúng và xây dựng một số chức năng kèm theo cho ứng dụng. Một số chức năng chính của Board điều khiển như sau: Xử lý chuỗi lệnh nhận được từ đó cho phép động cơ đóng mở cửa; hiển thị thông tin giao tiếp; gửi thông báo; mở cửa từ xa qua tin nhắn; còi chirp báo hiệu khi cửa đóng mở hay có người trong diện khả nghi.

Trong khối điều khiển còn thực hiện một số chức năng khác nhằm kết nối điều khiển động cơ và các cơ cấu chấp hành khác như báo hiệu, giao tiếp và hiển thị: như: Hiển thị một số thông tin, trạng thái của cửa lên LCD - Module Nokia 5510; thông báo SMS, hay mở cửa từ xa qua SMS bởi module SIM900A; sao chép dữ liệu khi cần thiết với module Micro SD; kết nối điều khiển động cơ Servo thông qua Jack cắm; còi chirp báo tín hiệu nhận dạng khuôn mặt còi sẽ kêu nếu là thành viên.. Khối cơ cấu chấp hành đóng mở cửa gồm mô hình cửa và Servo quay đóng mở cửa. Điều khiển góc quay động cơ bằng xung phương pháp điều chế độ rộng xung PWM. Phương pháp PWM giúp tạo ra những tín hiệu xung vuông mà ta có thể điều khiển được cả duty cycle và cả tần số nếu cần. Đối với việc điều khiển góc quay động cơ RC servo, tần số cần cố định ở 50Hz, thay đổi duty cycle sẽ thay đổi góc quay động cơ.

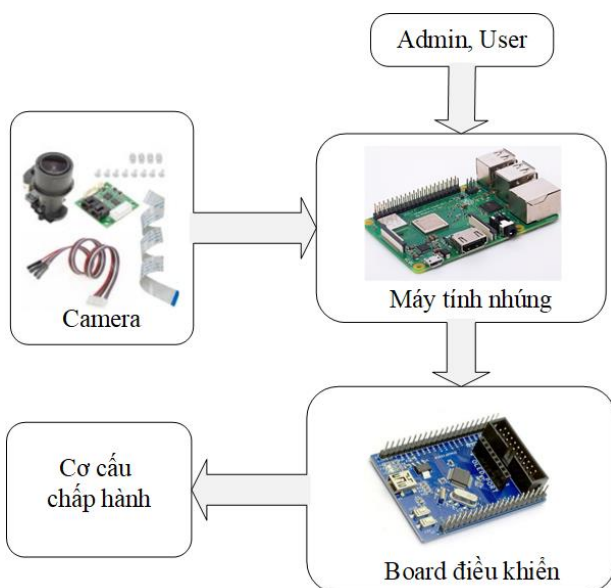


Figure 4: Các khối chính của hệ thống đóng mở cửa dựa trên nhận diện khuôn mặt.

3.2. Kết quả phát hiện khuôn mặt trực tiếp

Nhằm mục đích kiểm tra xem khuôn mặt đưa vào camera có thật sự là khuôn mặt người thực tế đang chụp hay không, người ta có thể thực hiện bằng cách kiểm tra chớp mắt hoặc sử dụng một số thiết bị phụ trợ khác. Trong nghiên cứu này, các tác giả đề xuất phương pháp xác định khuôn mặt trực tiếp dựa trên sự thay đổi cường độ sáng trên khuôn mặt nhờ nguyên lý về sự thay đổi của lưu lượng máu trong một chu kỳ tuần hoàn của tim.

Sau khi xác định được vùng khuôn mặt, ta tiến hành kiểm tra sự thay đổi mức xám trung bình của các hình ảnh liên tiếp trong chuỗi video thu thập được trong vòng 3 giây. Các giá trị mức xám tương ứng với cường độ sáng ở khuôn

mặt được tính cho cả ba kênh màu R, G, B. Với 30 ảnh thu thập trong một giây, thuật toán sẽ kiểm tra xem trong 3 giây đó, số lần thay đổi của cường độ sáng trung bình trên khuôn mặt có lớn hơn 15 hay không. Nếu đúng thì sẽ xác định là khuôn mặt thật. Điều này giúp loại trừ khả năng có đối tượng sử dụng hình ảnh giả mạo và di chuyển qua lại hình ảnh đó, với giả thiết là trong vòng 3 giây, khó có thể di chuyển hình ảnh giả bằng tay tới 15 lần.

Chúng tôi thực hiện thu thập dữ liệu, kiểm tra đánh giá việc nhận biết khuôn mặt trực tiếp của hệ thống trên 25 người khác nhau. Cụ thể, 25 hình khuôn mặt được in lên giấy và 25 người thật. Chúng tôi thực hiện ghi hình video có độ dài 3 giây cho mỗi khuôn mặt ở thời điểm khác nhau để kiểm tra sự thay đổi tín hiệu cường độ điểm ảnh trung bình. Kết quả cho thấy việc phát hiện người thật cho độ chính xác xấp xỉ 95%. Trong khi đó, nếu kiểm tra bằng việc kiểm tra chớp mắt [8, 9], độ chính xác đạt được chỉ ở mức 86% với thời gian 3 giây kể trên. Điều này cho thấy, trong một khoảng thời gian ngắn, có thể không xuất hiện sự chớp mắt của đối tượng trong khuôn hình, nhưng sự thay đổi về cường độ sáng trên khuôn mặt có thể được phát hiện ra. Kết quả thử nghiệm cho thấy là nếu ta ghi hình trong khoảng thời gian lâu hơn, ví dụ từ 5 đến 10 giây, ta có thể tính ra nhịp tim, và phát hiện khuôn mặt thật với độ chính xác cao hơn nữa. Tuy nhiên nếu tính nhịp tim thì thời gian cần phải chờ đợi để mở cửa sẽ lâu hơn vì cần phải tính cả thời gian ghi hình. Đồng thời cần thời gian xử lý ở bộ lọc thông trước khi tính ra các điểm cực đại trên tín hiệu hoặc biến đổi Fourier nhanh FFT. Do vậy thuật toán này chỉ nên áp dụng cho các ứng dụng mở cửa yêu cầu bảo mật cao, và cũng yêu cầu phần cứng với tốc độ xử lý nhanh hơn, dẫn đến giá thành cao hơn. Trong nghiên cứu này, với mục tiêu hướng tới là thiết kế hệ thống với giá thành thấp mà vẫn đáp ứng độ chính xác chấp nhận được, chúng tôi chỉ sử dụng sự thay đổi mức sáng dựa trên nguyên lý của thay đổi lưu lượng máu trong chu kỳ tuần hoàn. Kết quả cho thấy độ chính xác của thuật toán cao hơn so với phương pháp phổ biến là phát hiện chớp mắt.

3.3. Kết quả nhận diện khuôn mặt

Để phục vụ cho bài toán nhận dạng, trước hết cần xây dựng một thư viện ảnh huấn luyện. Thư viện huấn luyện hệ thống nhận diện gồm bộ hình ảnh positive và negative. Bộ positive chứa những hình ảnh khuôn mặt của 25 thành viên do chúng tôi tự thu thập. Các hình ảnh từ các thành viên này được chụp kèm theo với tên tương ứng của họ, trong đó mỗi người sẽ được chụp 20 ảnh khuôn mặt ở các thời điểm khác nhau trong ngày. Bộ hình ảnh negative chứa 400 ảnh của 40 người lạ, là các hình ảnh lựa chọn ngẫu nhiên từ các bộ cơ sở dữ liệu về khuôn mặt thường dùng trong lĩnh vực thị giác máy tính như Yale Face Database, và Google Facial Expression. Như vậy bộ ảnh dùng cho huấn luyện bao gồm 900 ảnh, trong đó có 500 ảnh positive là của thành viên, và 400 ảnh negative là của người lạ. Người quản lý cũng có thể dễ dàng bổ sung thêm các tập ảnh của thành viên mới vào tập huấn luyện. Nhờ đó bài toán có thể tăng độ tin cậy với số lượng ảnh tăng lên. Để kiểm nghiệm thuật toán, chúng tôi sử dụng 100 ảnh của thành viên và 100 ảnh không phải của thành viên.

Các bước thu thập ảnh từ camera, thuật toán xác định khuôn mặt cũng như nhận diện khuôn mặt được thực hiện trên máy tính nhúng Raspberry Pi, sử dụng phần mềm Py-

thon. Thuật toán nhận diện khuôn mặt trên thiết bị nhúng và ứng dụng trong đóng mở cửa được minh họa trên hình 5. Mới bắt đầu sau khi khởi động máy tính nhúng cùng Board điều khiển và các module SIM900A, MicroSD, LCD-Nokia5110, máy tính nhúng khởi tạo giao tiếp camera stream video, khởi tạo giao tiếp UART với Board mạch điều khiển thông qua đó điều khiển động cơ Servo để đóng mở cửa, đồng thời khởi tạo Server quản lý dữ liệu liên tục. Bằng việc xác định khuôn mặt có trong khung hình camera hay không mà ứng dụng đưa ra các bước xử lý tương ứng, nếu có khuôn mặt, việc nhận dạng danh tính người có trong khung hình được thực hiện. Thời gian xử lý cho một đối tượng trung bình khoảng 5 giây, trong đó lâu nhất là thời gian xác định khuôn mặt thật do phải thu thập một tập hình ảnh để tính sự thay đổi cường độ sáng. Bước nhận diện sẽ đưa ra danh tính của thành viên hay xác định là người lạ, từ đó cập nhật vào cơ sở dữ liệu đồng thời mở cửa rồi tự động đóng cửa sau 1 phút đi kèm gửi SMS thông báo.

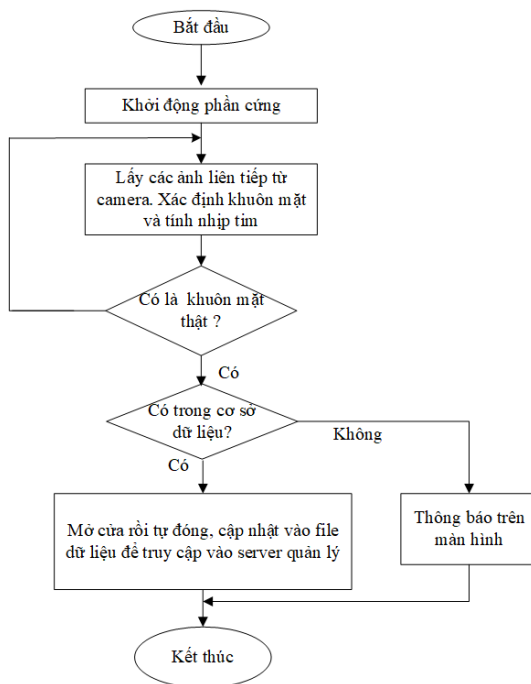


Figure 5: Lưu đồ thuật toán của bài toán nhận diện khuôn mặt trên máy tính nhúng cho ứng dụng mở cửa.

Với bài toán nhận diện khuôn mặt xem người đó có phải ở trong cơ sở dữ liệu để cho phép mở cửa hay không, độ chính xác được đánh giá trên cả phương diện positive và negative:

- Sai số positive là sai số của việc thành viên nhưng lại nhận diện là người lạ. Sai số trung bình của thuật toán là 17%.
- Sai số negative là sai số của việc thành viên này nhưng lại nhận diện là thành viên khác. Với kết quả đã thực nghiệm thì sai số này khoảng 4% , tức là nhận diện sai khoảng 1 người trong số 25 người.

4. Kết luận

Trong bài báo này, chúng tôi đã đưa ra một giải pháp cho bài toán nhận diện khuôn mặt trực tiếp và ứng dụng trong

đóng mở cửa tự động. Các hình ảnh được thu thập trực tiếp thông qua camera và xử lý trên máy tính nhúng. Các khuôn mặt sau khi đã được xác định sẽ được đưa qua bước kiểm tra xem đó là khuôn mặt trực tiếp hay là khuôn mặt giả từ ảnh/video. Việc kiểm tra này được thực hiện thông qua việc kiểm tra sự thay đổi của cường độ sáng của ảnh, giống như nguyên lý đo nhịp tim bằng phương pháp quang tử PPG. Cuối cùng các hình ảnh đã xác nhận là gương mặt trực tiếp sẽ được đưa vào nhận dạng xem có phải là người trong cơ sở dữ liệu cho phép được đi qua hệ thống hay không. Nếu người được nhận diện được xác định có trong cơ sở dữ liệu, hệ thống điều khiển sẽ cho phép mở cửa và lưu vào cơ sở dữ liệu. Hệ thống được xây dựng nhỏ gọn, giá thành thấp, có thể thích nghi cho một số ứng dụng khác như điểm danh, giám sát trong các hộ gia đình, hoặc là giám sát tài xế ngủ gật.

Acknowledgement

Nghiên cứu này được tài trợ bởi trường Đại học Bách khoa Hà Nội (HUST) trong đề tài mã số: T2020- PC-017

References

- [1] Daugman J. (2005) *How Iris Recognition Works*. Handbook of Image and Video Processing (Second Edition).
- [2] Palma D., Montessoro P. L., Giordano G., Blanchini F. (2019) *Biometric Palmprint Verification: A Dynamical System Approach*. IEEE Transactions on Systems, Man, and Cybernetics: Systems. vol. 49, pp. 2676 - 2687,
- [3] PA. V. , MJ. J. (2004) *Robust real-time face detection*. International Journal of Computer Vision volume vol. 57, pp. 137–154,
- [4] Georgescu D. (2011) *A Real-Time Face Recognition System Using Eigenfaces* Journal of Mobile, Embedded and Distributed Systems. vol. 4,
- [5] Wilson P. I. , Fernandez J. (2006) *Facial feature detection using Haar classifiers*. Journal of Computing Sciences in Colleges. vol. 21, pp. 127–133
- [6] Zhang K., Zhang L. ,Yang. M.H., "Real-time Compressive Tracking," in *European Conference on Computer Vision*, 2012, pp. 864-877
- [7] Zhang K., Zhang Z., Li Z. ,Qiao Y. (2016) *Joint face detection and alignment using multitask cascaded convolutional network*. IEEE Signal Processing Letters. vol. 23, pp. 1499 - 1503,
- [8] Kazemi V. , Sullivan J., "One millisecond face alignment with an ensemble of regression trees," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2014.
- [9] Soukupová T. , Cech J., "Real-time eye blink detection using facial landmarks," in *21st Computer Vision Winter Workshop*, 2016.
- [10] Poh M. Z., McDuff D. J. ,Picard R. W. (2010) Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics Express*. vol. 18, pp. 10762-10774,
- [11] Rahman H., Ahmed M. U., Begum S. ,Funk P., "Real Time Heart Rate Monitoring From Facial RGB Color Video Using Webcam," in *The 29th Annual Workshop of the Swedish Artificial Intelligence Society (SAIS)*, Malmö, Sweden, 2016.
- [12] Sirovitch L. , M. Kirby (1987) *Low-Dimensional Procedure for the Characterization of Human Faces*. Journal of the Optical Society of America A. vol. 2, pp. 519-524,
- [13] Turk M. , Pentland A. (1991) *Eigenfaces for Recognition*. Journal of Cognitive Neuroscience. vol. 3, pp. 71-86,