

NHẬN DẠNG MẶT NGƯỜI BẰNG MÁY HỌC CHUYÊN SÂU

ThS. Nguyễn Văn Danh¹, PGS.TS Phạm Thế Bảo²

Trường Cao đẳng Lý Tự Trọng TP.HCM¹, Trường Đại học Khoa học Tự nhiên TP.HCM²

Email: nvdanh81@gmail.com¹, ptbao@hcmus.edu.vn²

Tóm tắt: Trong cuộc cách mạng công nghiệp 4.0, trí tuệ nhân tạo được ứng dụng trong rất nhiều lĩnh vực khoa học và đời sống xã hội. Trong hội nghị về nhà trường thông minh. Muốn xây dựng một nhà trường thông minh, thì bài toán điểm danh sinh viên qua nhận dạng là không thể thiếu và đòi hỏi sự chính xác. Vì thế, bài toán nhận dạng mặt người đòi hỏi có độ chính xác cao. Trong bài báo này, chúng tôi đã đề xuất một phương pháp tiếp cận mới để nhận dạng khuôn mặt dựa trên hệ thống của mạng neural tích chập (Convolutional Neural Network - CNN). Với ảnh mặt người, chúng tôi áp dụng mô hình CNN trên từng thành phần của khuôn mặt, đó là thông tin yếu. Thông tin yếu là thông tin được trích xuất bởi các vùng của các thành phần của khuôn mặt, chẳng hạn như thành phần mắt trái, mắt phải, miệng và toàn mặt. Kết quả của hệ thống CNN là các bộ, các phần tử của mỗi thành phần mà có xác suất lớn nhất. Chúng tôi đã lấy kết quả cuối cùng bằng phương pháp kết hợp trễ và phương pháp bỏ phiếu. Chúng tôi đã thử nghiệm trên 02 cơ sở dữ liệu là ORL, CyberSoft. Thử nghiệm trên ORL và CyberSoft, chúng tôi phát hiện ra trọng số toàn cục của phương pháp bỏ phiếu. Sử dụng trọng số này, chúng tôi có độ chính xác cao nhất trên 02 cơ sở dữ liệu, như 99,375%, 99,5% trên cơ sở dữ liệu ORL, CyberSoft.

Keywords: Nhận dạng khuôn mặt, các thành phần khuôn mặt, mạng neural tích chập, phương pháp bỏ phiếu.

FACE RECOGNITION BY DEEP LEARNING

Nguyen Van Danh¹, Pham The Bao²

Ly Trong Technical College Ho Chi Minh City¹, VNUHCM – University of Science, Ho Chi Minh City²

Email: nvdanh81@gmail.com¹, ptbao@hcmus.edu.vn²

Abstract: In the industrial revolution 4.0, artificial intelligence is applied in many fields of science and social life. In the conference on smart school. To build a smart school, student identification is indispensable and requires accuracy. Therefore, the face recognition problem requires high accuracy. In this paper, we proposed a new approach for face recognition based on Convolutional Neural Network's system. A given face image, we applied Convolutional Neural Network (CNN) to each facial component region, which is weak information. The weak information are information were extracted by regions of face components, such as left eye region, right eye region, mouth region and the frontal face region. The results of CNN's system are sets, the elements of each set were the most probabilities of each classifier. We got the final result by late fusion with voting method. We experimented on 02 databases which are ORL, CyberSoft. Working on ORL and CyberSoft, I found out the global weights of voting method. Using this weights, we get the highest accuracy of 02 databases, which are 99.375%, 99.5% on ORL database, CyberSoft database.

Keywords: face recognition · facial components · multi-CNNs · late fusion · voting method.

1. Giới thiệu

Trong cuộc cách mạng công nghiệp 4.0, trí tuệ nhân tạo được ứng dụng trong rất nhiều lĩnh vực khoa học và đời sống xã hội. Trong hội nghị về nhà trường thông minh. Muốn xây dựng một nhà trường thông minh, thì bài toán điểm danh sinh viên qua nhận dạng là không thể thiếu và đòi hỏi sự chính xác. Vì thế, bài toán nhận dạng mặt người đòi hỏi có độ chính xác cao. Thực tế khi sinh viên vào trường và khi đi ngang qua hệ thống camera trong trường học hay trong lớp học thì hình ảnh thu được ở nhiều góc độ khác nhau,

có thể có nhiều góc nghiêng khác nhau, độ sáng khác nhau,...

Để thực hiện được vấn đề trên thì nhận dạng khuôn mặt một bài toán đòi hỏi có độ chính xác cao. Bài toán nhận dạng khuôn mặt vốn được nghiên cứu từ những năm 1970 và cho đến nay, rất nhiều nghiên cứu lần ứng dụng cho bài toán này đã ra đời. Nhiều phương pháp thực hiện với độ chính xác cao như PCA, LDA, Gabor và LBP, nhưng gặp khó khăn lớn khi ảnh mặt người bị tác động từ môi trường xung quanh, ví dụ như góc chụp, mức độ sáng, tối khi lấy

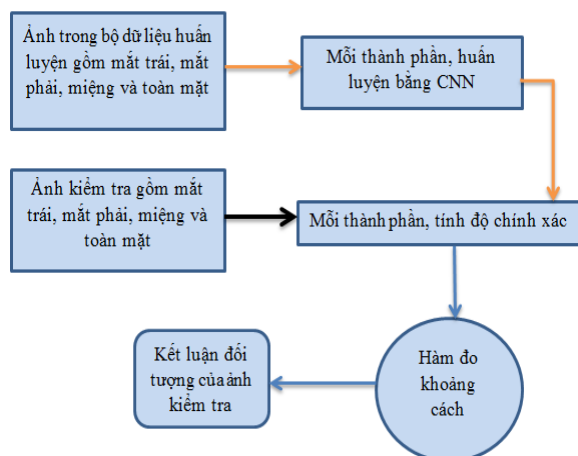
ảnh. Chất lượng ảnh đôi khi không được sắc nét do ảnh hưởng bởi nhiễu, phơi sáng. Ngoài ra, một người theo thời gian thì khuôn mặt sẽ biến đổi, do đó phần nào ảnh hưởng đến thông tin trên mặt như ria mép, râu, cũng như một số ảnh mặt người có đeo thêm phụ kiện như mắt kính, đôi khi vị trí của mắt kính vẫn khác nhau, như lệch xuống không đúng vị trí. Bên cạnh đó trạng thái của khuôn mặt cũng ảnh hưởng lớn đến vấn đề nhận dạng như vui, buồn, cười, há miệng, ngậm môi, nhắm mắt hay mở mắt to,...

Trong bài báo này, chúng tôi đề xuất phương pháp nhận dạng khuôn mặt người bằng máy học chuyên sâu cho từng thành phần khuôn mặt (như thành phần mắt trái, thành phần mắt phải, miệng và toàn mặt), rồi xây dựng bộ nhận dạng kết hợp từ các CNN thành phần.

Với CNN, khi nhận dạng với dữ liệu trích đặc trưng tự động là ảnh toàn mặt, thì có thể nhiều thông tin không "nhìn thấy", nhưng nếu tách riêng thì có thể các đặc trưng này sẽ được "nhìn thấy". Vì thế chúng tôi quyết định xây dựng bộ nhận dạng kết hợp từ các CNN thành phần. Kết quả chúng tôi thực nghiệm trên bộ dữ liệu ORL là 99.375% cao hơn so với các báo cáo trước đây như [1], [6], [7], [8] và [9] lần lượt với kết quả 86.070%, 99.250%, 94.170%, 97.500% và 91.000%. Khi chúng tôi thực nghiệm trên bộ dữ liệu CyberSoft đạt 99.5% cao hơn so với kết quả thực nghiệm của phương pháp [6] với kết quả là 90.0%.

2. Phương pháp thực hiện

Tóm tắt thuật toán như Hình 2.1, từ các ảnh ban đầu trong bộ dữ liệu và ảnh kiểm tra, các dữ liệu này đã được xử lý để tách dữ liệu cho thành phần có trên khuôn mặt như thành phần mắt trái, thành phần mắt phải, thành phần miệng, và hình ảnh toàn mặt.



Hình 2.1. Tóm tắt quy trình tổng quát của đề tài

Trong bài báo này, chúng tôi tiến hành xây dựng mô hình CNN cho các thành phần như : Thành phần

dữ liệu ảnh mắt trái, mắt phải, miệng và toàn mặt. Từ đó, chúng tôi sẽ hiện thực để tính độ chính xác nhận dạng trên những thành phần như:

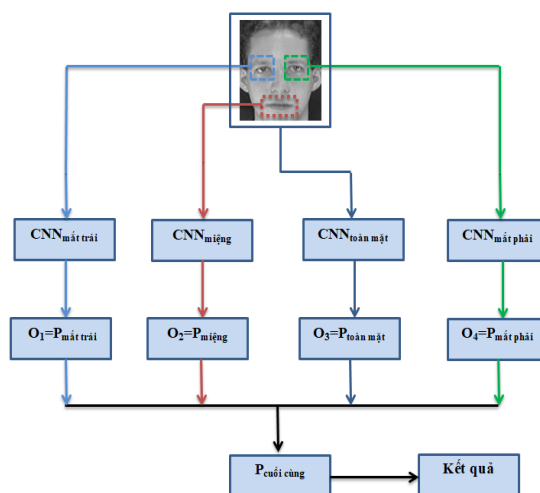
- Độ chính xác trên mỗi thành phần dữ liệu: thành phần mắt trái, thành phần mắt phải, thành phần miệng và thành phần toàn mặt
- Độ chính xác khi kết hợp cả 3 thành phần: thành phần mắt trái, thành phần mắt phải, thành phần miệng
- Độ chính xác khi kết hợp cả 4 thành phần: thành phần mắt trái, thành phần mắt phải, thành phần miệng và toàn mặt.

2.1 Mô hình tổng quát

Trong bài báo này, chúng tôi đề xuất mô hình nhận dạng khuôn mặt người, Hình 2.2. Mô hình này được tích hợp từ 4 mô hình CNN tương ứng cho từng thành phần: thành phần mắt trái, thành phần mắt phải, thành phần miệng và toàn mặt.

Sau khi huấn luyện mô hình CNN cho từng mô hình CNN thành phần, với mỗi mô hình CNN_i (với i=[1..4]), chúng tôi sử dụng m_j (với j= [1..5]) phần tử nhận dạng cho khả năng cao nhất CNN_i gọi là tập các phần tử $O_i = \{P_{i,m_j}\}_{m_j}$.

Sau đó áp dụng hệ thống phân loại tích hợp từ nhiều phân loại khác nhau để tính ra độ chính xác để nhận dạng một đối tượng

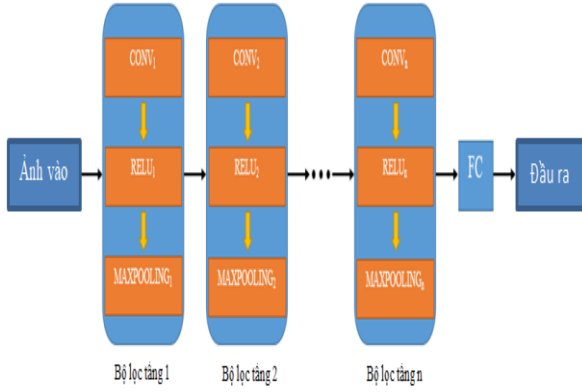


Hình 2.2. Mô hình tổng quát

2.2 Kiến trúc CNN tổng quát

Trong bài báo này, chúng tôi lựa chọn kiến trúc CNN cho từng phần, Hình 2.3. Với thứ tự trong một tầng đi từ giai đoạn tích chập (CONV), rồi sử dụng hàm kích hoạt ReLU, và cuối cùng là Pooling. Các tầng còn lại cũng lặp lại cấu trúc như vậy .

Trong bài báo này, quá trình thực nghiệm chúng tôi sử dụng mô hình CNN, Hình 2.3. Gồm 3 tầng, trong từng tầng tôi sử dụng đầy đủ 3 giai đoạn: giai đoạn một là tính tích chập (CONV), giai đoạn 2 thực hiện hàm kích hoạt(RELU) và giai đoạn cuối trong tầng là thực hiện Pooling, giai đoạn pooling được sử dụng ở đây là max pooling.



Hình 2.3. Mô hình tổng quát

Cuối cùng, lớp Fully-Connected (FC) cho ra vector đặc trưng của ảnh, dùng để tính điểm các đối tượng.

2.3 Xây dựng hệ thống phân loại tích hợp từ nhiều phân loại khác nhau

Vì lý do thông tin ở các thành phần ít và có thể có thành phần không có thông tin nên chúng tôi xây dựng bộ phân loại kết hợp từ n kết quả của n bộ phân loại trước đó (cụ thể trong bài báo này tôi có 4 thành phần là mắt trái, mắt phải, miệng và toàn mặt) dựa vào trọng số xác suất của từng bộ phân loại. Mỗi kết quả của một bộ phân loại xem như là một biến ngẫu nhiên.

Cho ảnh kiểm tra, giả sử chúng tôi chia mặt ra thành n thành phần gồm thành phần mắt trái, mắt phải, miệng và thành phần toàn mặt, khi đó với mỗi thành phần i , chúng tôi chọn k đối tượng có thành phần i trên mặt có xác suất giống với thành phần i trên ảnh kiểm tra nhất, sau đó chúng tôi tính điểm k đối tượng đấy. Sau n thành phần, đối tượng nào có điểm cao nhất, chúng tôi kết luận đó chính là đối tượng thuộc ảnh kiểm tra.

Xét đối tượng a , hàm tính xác suất a có khả năng phân loại đúng là $P_{cuối cùng}$ của đối tượng a là

$$P_{cuối cùng} = \operatorname{argmax} \sum_{i=1}^n \alpha_i \times P_i(\text{Object}_k) \quad (1)$$

Với $B = \bigcup_1^n O_i \quad k \in B$

Trong đó:

α_i : Hệ số của thành phần i .

$P_i(\text{Object}_k)$: Xác suất thành phần i của đối tượng k giống với thành phần i trong ảnh kiểm tra.

n : Số thành phần sử dụng để nhận dạng.

Chúng tôi xác định hệ số thành phần bằng cách sau khi huấn luyện từng thành phần i trên ảnh huấn luyện, chúng tôi đưa thành phần i trong ảnh kiểm tra vào để xác định độ chính xác β_i , sau đó, chúng tôi tính tổng độ chính xác. Khi đó, hệ số α_i bằng β_i chia cho tổng độ chính xác. Chúng tôi thu được giá trị $P_i(\text{Object}_k)$ trong quá trình huấn luyện trên từng thành phần.

$$\alpha_i = \frac{\beta_i}{\sum_{j=0}^n \beta_j} \quad (2)$$

Sau khi kiểm tra trên n thành phần, đối tượng nào có điểm cao nhất thì chúng tôi kết luận ảnh kiểm tra chính là của đối tượng đó. Dữ liệu trên thực nghiệm sẽ xác định được cụ thể hệ số α_i .

2.4 Xây dựng hệ số quan trọng α đại diện cho nhiều bộ dữ liệu

Sau khi tính được hệ số quan trọng cho từng bộ dữ liệu, chúng tôi đề xuất công thức tính hệ số quan trọng α chung đại diện cho các bộ dữ liệu theo công thức (3).

$$\alpha_{chung} = \frac{1}{k} \sum_{i=1}^k \alpha_i \quad (3)$$

Trong đó:

k là số bộ dữ liệu.

$\alpha_i \in \mathbb{R}^N$, $i = \overline{1..n}$ là các bộ trọng số của một bộ dữ liệu.

3. Thử nghiệm

Chúng tôi thực nghiệm trên máy sử dụng hệ điều hành Windows 10 Pro bản 64 bit với RAM 8GB, Intel(R) Core(TM) i7-3632QM CPU @ 2.20GHz. Trong quá trình huấn luyện, chúng tôi sử dụng Python kết hợp với thư viện mã nguồn mở TensorFlow do Google phát triển, chạy trên nền Jupyter Notebook.

Trong bài báo này, chúng tôi thực nghiệm trên 2 bộ dữ liệu ORL và CyberSoft. Trong từng bộ dữ liệu, tôi chia ra làm 4 thành phần, thành phần mắt trái, mắt phải, miệng và toàn mặt. Với từng bộ dữ liệu thành phần, tôi lấy 80% hình ảnh để huấn luyện, sau khi huấn luyện, bộ dữ liệu kiểm tra gồm 20% của bộ dữ liệu chưa được huấn luyện và 25% của 80% bộ dữ liệu đã được huấn luyện cho từng thành phần riêng.

3.1 Bộ dữ liệu ORL

Bộ dữ liệu ORL được thu thập từ tháng 4 năm 1992 đến tháng 4 năm 1994, của Đại học Cambridge.

Trong bộ dữ liệu ORL, gồm có 1600 ảnh, bao gồm 4 thành phần như thành phần mắt trái, thành phần mắt phải, miệng và toàn mặt. Trong mỗi thành phần gồm 400 ảnh, cho 40 đối tượng khác nhau, mỗi đối tượng có 10 ảnh. Với 256 mức độ xám của một điểm ảnh.



Hình 3. 1 Một số hình trên bộ dữ liệu ORL

Bảng 1. Kích thước mỗi thành phần trên bộ dữ liệu ORL

Thành phần	Kích thước	Hình đại diện
Thành phần mắt trái	32×32	
Thành phần mắt phải	32×32	
Thành phần miệng	40×24	
Toàn mặt	112×92	

3.2 Bộ dữ liệu Cybersoft

Bộ dữ liệu này được thu thập bởi công ty Global CyberSoft. Với bộ dữ liệu này, gồm 3200 ảnh cho 4 thành phần. Mỗi thành phần có 800 ảnh cho 40 đối tượng khác nhau. Mỗi đối tượng có 20 ảnh.



Hình 3. 2 Một số hình trên bộ dữ liệu CyberSoft

Bảng 2. Kích thước của mỗi thành phần trên bộ dữ liệu CyberSoft

Thành phần	Kích thước	Hình đại diện
Thành phần mắt trái	32×32	
Thành phần mắt phải	32×32	
Thành phần miệng	40×26	
Toàn mặt	80×80	

3.3 Tính đa dạng của 2 bộ dữ liệu

Với 2 bộ dữ liệu ORL và CYBERSOFT là các hình ảnh được chụp trực diện, theo phương thẳng đứng, với nhiều góc độ chụp khác nhau, với nhiều trạng thái khác nhau, ở những thời điểm sử dụng các phụ kiện trên khuôn mặt khác nhau. Cả hai bộ dữ liệu này đa dạng, cụ thể:

Ảnh khuôn mặt người có tính đa dạng, ví dụ như có lúc đeo kính, có lúc không đeo kính, xoay sang phải hay sang trái một góc không đáng kể, đảm bảo không mất thông tin trên khuôn mặt như mắt trái, mắt phải và miệng.

Cả hai bộ dữ liệu, các ảnh được chụp khuôn mặt chính diện, có một số ảnh nghiêng một góc không đáng kể để đảm bảo không mất các thành phần trên khuôn mặt. Đồng thời ảnh của bộ dữ liệu được chụp ở nhiều trạng thái khác nhau như cười không cười. Lúc thì cười với môi ngậm, lúc cười với 2 ở trạng thái hở răng,....

Các góc nhìn của đối tượng khác nhau, như nhìn lên, nhìn xuống, nhìn sang trái, sang phải, có thể kết hợp nhiều hình thái như vừa cười vừa nheo mắt, hay mở mắt to, hay nheo mắt lại. Chịu tác động bởi ngoại cảnh, ảnh bị mờ. Thiết bị lấy ảnh khác nhau, như Samsung, Iphone, Canon,... nên độ phân giải khác nhau, chất lượng màu sắc, độ sáng khác nhau.

3.4 Chọn hệ số cho hệ thống CNN

Chúng tôi sử dụng kiến trúc CNN 3 tầng. Trong mỗi tầng có đầy đủ 3 giai đoạn như CONV, RELU và POOL. Bảng 4.3, sẽ chỉ ra các hệ số ở giai đoạn nhân tích chập (CONV) trên từng tầng, cũng như kích thước bộ lọc, số bộ lọc, bước nhảy của bộ lọc trên cửa sổ trượt.

Trong giai đoạn Pooling của 3 tầng tôi đều dùng bộ lọc có kích thước 2×2 , nhưng số bộ lọc trong từng

tầng thì bằng số bộ lọc của giai đoạn CONV (Bảng 3), cả 3 tầng bước nhảy của bộ lọc này đều là 2.

Bảng 3. Kích thước của bộ lọc CONV của các tầng

Tầng	Kích thước	Số bộ lọc	Bước nhảy
CONV ₁	5 × 5	16	1
CONV ₂	4 × 4	32	1
CONV ₃	3 × 3	48	2

3.5 Xây dựng bộ hệ số quan trọng của hệ thống tích hợp

Với bộ dữ liệu ORL: Chúng tôi huấn luyện và kiểm tra trên 4 thành phần gồm mắt trái, mắt phải, miệng và toàn mặt, thu được độ chính xác trên từng phần cũng như là xác suất xây dựng bộ hệ số quan trọng cho từng thành phần, Bảng 4.

Bảng 4. Độ chính xác trên bộ dữ liệu ORL

Thành phần	Độ chính xác	Tên trọng số
Thành phần mắt trái	87.500%	β_1^{ORL}
Thành phần mắt phải	83.125%	β_2^{ORL}
Thành phần miệng	88.125%	β_3^{ORL}
Toàn mặt	98.125%	β_4^{ORL}

Với dữ liệu trên Bảng 4, ta chỉ lấy thông tin 3 thành phần gồm mắt trái, mắt phải và thành phần miệng. Áp dụng công thức (1), chúng tôi tính được hệ số quan trọng α_{13} cho 3 thành phần trên, Bảng 5.

Bảng 5. Trọng số trên 3 thành của bộ dữ liệu ORL

Thành phần	Tên	Trọng số
Thành phần mắt trái	$\alpha_{1,ORL}^3$	140/414
Thành phần mắt phải	$\alpha_{2,ORL}^3$	133/414
Thành phần miệng	$\alpha_{3,ORL}^3$	141/414

Với dữ liệu trên Bảng 4, chúng tôi lấy thông tin trên tất cả 4 phần: mắt trái, mắt phải, miệng và toàn mặt. Áp dụng công thức (1), chúng tôi tính được hệ số quan trọng α_{14} cho 4 thành phần trên, Bảng 6.

Bảng 6. Trọng số trên 4 thành phần của bộ dữ liệu ORL

Thành phần	Tên	Trọng số
Thành phần mắt trái	$\alpha_{1,ORL}^4$	140/571
Thành phần mắt phải	$\alpha_{2,ORL}^4$	133/571
Thành phần miệng	$\alpha_{3,ORL}^4$	141/571
Toàn mặt	$\alpha_{4,ORL}^4$	157/571

Với bộ dữ liệu CYBERSOFT: Chúng tôi cũng huấn luyện và kiểm tra trên 4 thành phần gồm thành phần mắt trái, thành phần mắt phải, thành phần miệng và toàn mặt, thu được độ chính xác trên từng phần cũng như là xác suất xây dựng bộ hệ số quan trọng cho từng thành phần, Bảng 7.

Bảng 7. Độ chính xác trên bộ dữ liệu CyberSoft

Thành phần	Độ chính xác	Tên trọng số
Thành phần mắt trái	99.0%	β_1^{CS}
Thành phần mắt phải	98.5%	β_2^{CS}
Thành phần miệng	94.5%	β_3^{CS}
Toàn mặt	99.0%	β_4^{CS}

Với dữ liệu trên Bảng 7, ta chỉ lấy thông tin 3 thành phần gồm mắt trái, mắt phải và thành phần miệng. Áp dụng công thức (1), chúng tôi tính được hệ số quan trọng α_{23} cho 3 thành phần trên, Bảng 8.

Bảng 8. Trọng số trên 3 thành của bộ dữ liệu CyberSoft

Thành phần	Tên	Trọng số
Thành phần mắt trái	$\alpha_{1,CS}^3$	198/584
Thành phần mắt phải	$\alpha_{2,CS}^3$	197/584
Thành phần miệng	$\alpha_{3,CS}^3$	189/584

Với dữ liệu trên Bảng 7, ta lấy thông tin 4 thành phần gồm thành phần mắt trái, thành phần mắt phải, thành phần miệng và toàn mặt. Áp dụng công thức (1), chúng tôi tính được hệ số quan trọng α_{24} cho 4 thành phần trên, Bảng 9.

Bảng 9. Trọng số trên 4 thành phần của bộ dữ liệu CyberSoft

Thành phần	Tên	Trọng số
Thành phần mắt trái	$\alpha_{1,CS}^4$	198/782
Thành phần mắt phải	$\alpha_{2,CS}^4$	197/782
Thành phần miệng	$\alpha_{3,CS}^4$	189/782
Toàn mặt	$\alpha_{4,CS}^4$	198/782

Từ dữ liệu hệ số quan trọng α_{13} cho 3 thành phần trên bộ dữ liệu ORL, Bảng 5 và hệ số quan trọng α_{23} cho 3 thành phần trên bộ dữ liệu CYBERSOFT, Bảng 8. Áp dụng công thức (3), chúng tôi tính hệ số quan trọng chung với 3 thành phần: thành phần mắt trái, thành phần mắt phải và thành phần miệng trên cả 2 bộ dữ liệu ORL và CYBERSOFT, Bảng 10.

Bảng 10. Hệ số quan trọng α_{33} chung cho 3 thành phần

Thành phần	Tên	Trọng số
Thành phần mắt trái	$\alpha_{3,global}^1$	3386/10000
Thành phần mắt phải	$\alpha_{3,global}^2$	3293/10000
Thành phần miệng	$\alpha_{3,global}^3$	3321/10000

Từ dữ liệu hệ số quan trọng α_{14} cho 4 thành phần trên bộ dữ liệu ORL, Bảng 6 và hệ số quan trọng α_{24} cho 4 thành phần trên bộ dữ liệu CYBERSOFT, Bảng 9. Áp dụng công thức (3), chúng tôi tính hệ số quan trọng chung với 4 thành phần trên 2 bộ dữ liệu, Bảng 11.

Bảng 11. Hệ số quan trọng α_{34} chung cho 4 thành phần

Thành phần	Tên	Trọng số
Thành phần mắt trái	$\alpha_{4,global}^1$	2492/10000
Thành phần mắt phải	$\alpha_{4,global}^2$	2424/10000
Thành phần miệng	$\alpha_{4,global}^3$	2443/10000
Toàn mặt	$\alpha_{4,global}^4$	2641/10000

3.6 Kết quả thực nghiệm

Chúng tôi tiến hành thực nghiệm trên từng hệ số riêng, và hệ số chung, lần lượt thu được kết quả về độ chính xác trên từng bộ dữ liệu như sau:

Độ chính xác với hệ số riêng trong Bảng 12.

Bảng 12. Độ chính xác khi thực hiện trên bộ dữ liệu ORL, CyberSoft, sử dụng hệ số riêng

Bộ dữ liệu	3 thành phần	4 thành phần
ORL	95.000%	99.375%
CyberSoft	99.500%	99.500%

Độ chính xác khi sử dụng hệ số chung của 2 bộ dữ liệu ORL và CyberSoft

Bảng 13. Độ chính xác khi thực hiện trên bộ dữ liệu ORL, CyberSoft, sử dụng hệ số chung

	3 thành phần	4 thành phần
ORL	94.375%	99.375%
CyberSoft	99.500%	99.500%

Kết quả chúng tôi đạt được trong bài báo này khi kết hợp các thành phần lại, chúng tôi có kết quả cao hơn so với các phương pháp đã làm trước đây. Cụ thể trong bài báo này tôi sử dụng 5 kết quả của các tác giả thực hiện trên bài toán nhận dạng mặt người nhưng sử dụng các phương pháp khác. Cụ thể có 4 phương pháp thực hiện trên bộ dữ liệu ORL và 1 phương pháp thực hiện trên cả 2 bộ dữ liệu ORL và CYBERSOFT, Bảng 14

Bảng 14. So sánh với kết quả của các bài đã làm trước đây

Thuật toán	ORL	CyberSoft
[1]	86.070%	-
[6]	99.250%	90.0%
[7]	94.170%	-
[8]	97.500%	-
[9]	91.000%	-
Chúng tôi (3 thành phần)	94.375%	99.5%
Chúng tôi (4 thành phần)	99.375%	99.5%

4. Kết luận

Trong bài báo này, chúng tôi đã đề xuất một phương pháp nhận dạng khuôn mặt dựa trên việc kết hợp giữa các thành phần yếu để đạt được phân loại mạnh và sử dụng phương pháp kết hợp muộn để tăng hiệu suất nhận dạng. Với một dữ liệu khuôn mặt, tôi chia hình ảnh trong mỗi đối tượng thành 4 thành phần: thành phần mắt trái, mắt phải, miệng và toàn mặt. Tôi không sử dụng thành phần mũi, vì đối với các hình có góc chụp hơi nghiêng trái hay nghiêng phải, cúi đầu hay ngược lên sẽ làm mất thông tin của mũi. Chúng tôi đã sử dụng mô hình CNN dựa trên kiến trúc LeNet-5 để huấn luyện và kiểm tra trên từng thành phần. Cuối cùng kết hợp các thành phần này lại để tăng hiệu suất nhận dạng. Chúng tôi đã thử nghiệm trên cơ sở dữ liệu ORL, CyberSoft, Georgia Tech. Kết quả chúng tôi đạt được 99.375% trên bộ dữ liệu ORL, 99.5% trên CyberSoft khi sử dụng tất cả 4 thành phần: Mắt trái, mắt phải, miệng và toàn mặt.

Tài liệu tham khảo

- [1] Phạm Thế Bảo và Trịnh Tấn Đạt, "Dùng đặc trưng gabor kết hợp adaboost và k-means trong bài toán nhận dạng mặt người", Tạp chí Khoa học ĐHSP Thành phố Hồ Chí Minh, Số 43, 2013.
- [2] Võ Phúc Nguyên, "Nhận dạng ảnh mặt người sử dụng mạng nơron", Tạp chí Khoa học & Công nghệ, Số 64, trang 52-57, 2014.
- [3] Võ Hoàng Trọng, Nguyễn Thanh Thủy, "Nhận dạng mặt người dựa trên thông tin mặt người không đầy đủ", Luận văn tốt nghiệp đại học, Đại học Khoa học Tự nhiên, Thành phố Hồ Chí Minh, 2017.
- [4] T. Ahonen, A. Hadid and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," IEEE transactions on pattern analysis and machine

- intelligence, vol. 28, no. 12, pp. 2037-2041, 2006.
- [5] Y. Taigman, M. Yang, M. A. Ranzato and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1701-1708, 2014.
 - [6] H. V. Truong, V. Ngo and P. T. Bao, "Face recognition based on hybrid Contourlet features with Bayesian network," International Conference on Video and Image Processing (ICVIP'17), 2017.
 - [7] Y. Zhang, D. Zhao, J. Sun, G. Zou and W. Li, "Adaptive convolutional neural network and its application in face recognition," Neural Processing Letters, vol. 43, no. 2, 2016.
 - [8] Y. Gan, T. Yang and C. He, "A deep graph embedding network model for face recognition," Signal Processing (ICSP), 2014 12th International Conference on, pp. 1268-1271, 2014.
 - [9] W. Wang, J. Yang, J. Xiao, S. Li and D. Zhou, "Face recognition based on deep learning," International Conference on Human Centered Computing, pp. 812-820, 2014.
 - [10] Google Brain Team, "TensorFlow: An open-source software library for Machine Intelligence," [Online]. Available: <https://www.tensorflow.org/>.