

PHÁT HIỆN VÀ ĐỊNH DANH KHUÔN MẶT

I. Giới thiệu

Con người rất giỏi trong việc nhận diện khuôn mặt và các hình mẫu phức tạp. Ngay cả khi một thời gian trôi qua vẫn không ảnh hưởng đến khả năng này và con người muốn máy tính trở nên mạnh mẽ như họ trong việc nhận diện khuôn mặt để hỗ trợ con người.

A. Mục tiêu

Tìm hiểu và nghiên cứu các phương pháp nhận dạng khuôn mặt. Kết quả nhận dạng được dùng làm thông tin để định danh người có là những người quen biết đối với bản thân hay không.

Tìm hiểu phương pháp trích xuất đặc trưng khuôn mặt và lưu để thử nghiệm.

Xây dựng phần mềm và thực hiện hệ thống thử nghiệm.

B. Bài toán máy học

Xác định các vị trí và các kích thước của khuôn mặt trong bất kì (ảnh kĩ thuật số, hình ảnh trong các bài báo, camera an ninh...) và xác định đó có phải là người quen của ta hay là không.

C. Ứng dụng

- Mở cửa nhà, lớp học, văn phòng...
- Tra cứu thông tin tội phạm
- Giám sát bằng camera để phát hiện tội phạm tại khu vực công cộng
- v/v...

II. Dữ liệu

A. Giới thiệu tập dữ liệu có sẵn

- 1) *Haarcascade_frontalface_default.xml*. (Preda, 2020)



H.1. Data training haar-like feature

Phát hiện mặt người là bài toán cơ bản được xây dựng từ nhiều năm nay, có nhiều phương pháp được đưa ra như sử dụng template matching, neuron network... Cho tới nay bài toán này hầu như được giải quyết dựa trên phương pháp sử dụng các đặc trưng haar-like.

Phương pháp haar-like được cho là đơn giản và kết quả phát hiện là tương đối cao, lên tới 98%. Các hãng sản xuất máy ảnh như Canon, Samsung... cũng đã tích hợp nó vào trong các sản phẩm của mình.

Vì sự phổ biến hiện nay, phương pháp này được train sẵn để có thể được sử dụng dễ dàng và rộng rãi hơn. Kết quả của phương pháp haar-like được đúc kết thành file "*Haarcascade_frontalface_default.xml*" để tiện cho việc tái sử dụng.

- 2) *shape_predictor_68_face_landmarks.dat* (Masa, 2020)

Ở những phương pháp nhận diện khuôn mặt trước đây, chúng ta gặp phải khó khăn khi khuôn mặt chụp với góc độ khác nhau sẽ khác nhau đối với máy tính.

Để giải quyết vấn đề này, phải biến đổi sao cho mắt mũi môi ở cùng một vị trí trong ảnh. Việc này giúp nhận diện ảnh tốt hơn.

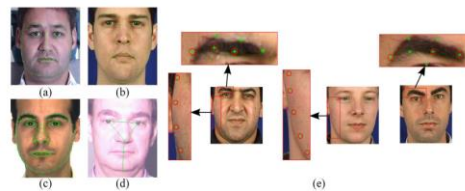


H-2. Chuyển mắt mũi môi về trung tâm

Sau khi tinh chỉnh thì tiến hành trích xuất dữ liệu và tiến hành training. Bằng cách sử dụng thuật toán của Vahid Kazemi và Josephine được biết đến với tên Face landmark estimation.

Tương tự như đã đề cập như tập dữ liệu trước, để được sử dụng rộng rãi.

"*shape_predictor_68_face_landmarks.dat*" dựa trên bộ dữ liệu ảnh chuẩn ibug 300-W.



H-3. Xác định 68 điểm mốc trên khuôn mặt

- 3) *dlib_face_recognition_resnet_model_v1.dat* (King, 2020)

Việc nhận dạng rất dễ dàng so với con người đơn giản như màu mắt, long mày dày, mũi cao, mặt tròn... nhưng với máy tính thì hoàn toàn khác vì nó chỉ có thể phân biệt được từng điểm ảnh.

Mạng CNN – Convolution Neural Network làm tốt hơn trong việc tìm ra yếu tố đo lường quan trọng.

Model `dlib_face_recognition_resnet_model_v1` sẽ hỗ trợ mạnh mẽ việc nhận dạng này.

B. Xây dựng tập dữ liệu

Với mục tiêu định danh người có là những người quen biết đối với bản thân hay không. Tiến hành thu thập ảnh khuôn mặt của người quen.



H-4. Data thu thập

C. Phân tích sơ bộ tập dữ liệu

Bộ dữ liệu chúng ta thu thập bao gồm 19 người và tiến hành gán nhãn:

- Võ Đại Dương
- Từ Thanh Dương
- Phạm Trung Hoà
- Trương Nguyễn Trung Anh
- Trần Tuấn Anh
- Tăng Ngọc Mi
- Nguyễn Minh Tâm
- Phạm Quang Hải
- Phan Phước Đạt
- Nguyễn Tiến Bảo
- Nguyễn Thiên Đạt
- Nguyễn Đức Hoàn
- Nguyễn Chí Dũng
- Nguyễn Ngọc Hải
- Ngô Viết Hiếu
- Long Hải
- Kim Anh Đức
- Di Đan
- Bùi Huy Hoàng

III. Phương pháp máy học

Thuộc phương pháp học có giám sát (Supervised learnig) sử dụng cho bài toán phân lớp (Classification).

Với bộ (input, outcome) tương ứng (ảnh khuôn mặt, tên người)

Từ dữ liệu thu thập được từ 19 người, phân thành 20 lớp tương ứng – 19 người từ data và không thể định danh được.



H-5. Mô hình bài toán định danh khuôn mặt

A. Các phương pháp máy học

Đầu tiên, để đi vào các phương pháp được sử dụng trong bài toán này, cần phải biết được ảnh trong máy tính được hiểu như thế nào.

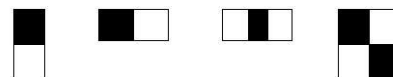
- Hệ màu RGB: Với mỗi bộ 3 số r, g, b nguyên trong khoảng $[0, 255]$ sẽ cho ra một màu khác nhau. Do có 256 cách chọn r, 256 cách chọn màu g, 256 cách chọn b vậy tổng số màu có thể tạo ra bằng hệ màu RGB là $256^3 = 16777216$ màu.
- Ảnh màu: sẽ được biểu diễn dưới dạng tensor 3 chiều do có 3 ma trận (channel) màu red, green, blue kích thước $H \times W \times 3$.

1) Phương pháp xác định khuôn mặt của Violas & John (Haar-like) (HaiHa, 2019)

Như đã giới thiệu ở phần trước, phát hiện mặt người là bài toán cơ bản được xây dựng từ nhiều năm nay và được sử dụng rộng rãi trên thế giới. Phương pháp haar-like được cho là đơn giản và kết quả phát hiện là tương đối cao, lên tới 98%. Ở phần này ta sẽ đi sâu vào thuật toán này.

- Các đặc trưng haar-like:

Là những hình chữ nhật được phân thành các vùng đen trắng. Violas & John công bố gồm 4 đặc trưng cơ bản để xác định khuôn mặt. Mỗi đặc trưng là sự kết hợp của hai hay ba hình chữ nhật đen trắng như trong hình sau.



H-6. Các đặc trưng haar-like

Viola & Joines đưa ra một khái niệm gọi là Integral Image, là một mảng 2 chiều với kích thước

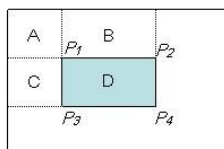
bằng với kích thước của ảnh cần tính đặc trưng Haar-Like, với mỗi phần tử của mảng này được tính bằng cách tính tổng của điểm ảnh phía trên (dòng-1) và bên trái (cột-1) của nó.



Công thức tính Integral Image

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

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

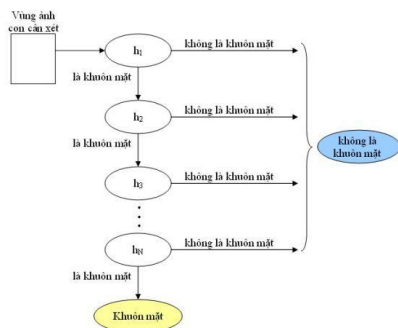


$$D = A + B + C + D - (A + B) - (A + C) + A$$

Với $A + B + C + D$ chính là giá trị tại điểm P_4 trên Integral Image, tương tự như vậy $A+B$ là giá trị tại điểm P_2 , $A+C$ là giá trị tại điểm P_3 , và A là giá trị tại điểm P_1 .

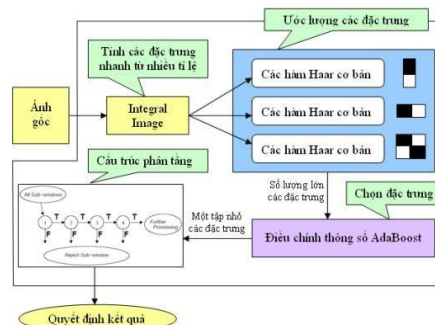
- AdaBoost:

Adaboost cũng hoạt động trên nguyên tắc kết hợp tuyến tính các weak classifiers để hình thành một trong các classifiers. Viola và Jones dùng AdaBoost kết hợp các bộ phân loại yếu sử dụng các đặc trưng Haar-like theo mô hình phân tầng (cascade) như sau:



H-7.mô hình phân tầng (cascade)

- Sơ đồ xác định khuôn mặt:



H-8. Sơ đồ xác định khuôn mặt

2) Phương pháp nhận diện khuôn mặt bằng Facial Landmark (Quân, 2018)

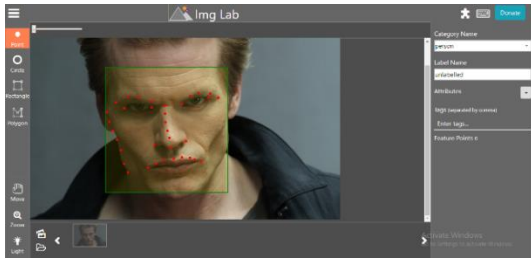
Để giải quyết vấn đề khi khuôn mặt chụp với góc độ khác nhau sẽ khác nhau đối với máy tính. Với ý tưởng xác định 68 điểm mốc trên khuôn mặt (như trong châm cứu) tồn tại trên mỗi khuôn mặt bằng cách huấn luyện một thuật toán máy học.



H-9.Sơ đồ 68 điểm mốc trên khuôn mặt

Khi xác định được điểm mốc trên khuôn mặt đã biết và khuôn mặt cần nhận dạng. Có thể tính được độ cao, rộng của mắt, mũi, và miệng, khoảng cách giữa hai mắt của khuôn mặt đã biết và khuôn mặt cần nhận dạng và so sánh chúng để tìm ra điểm tương đồng.

Để xác định được 68 điểm mốc trên khuôn mặt, Davis King đã sử dụng bộ dữ liệu ảnh chuẩn ibug 300-W và cho mạng CNN để huấn luyện cho ra file huấn luyện shape_predictor_68_face_landmarks.dat. Để tạo ra 68 điểm cột mốc trên khuôn mặt Davis King đã dùng Imglab để định vị.

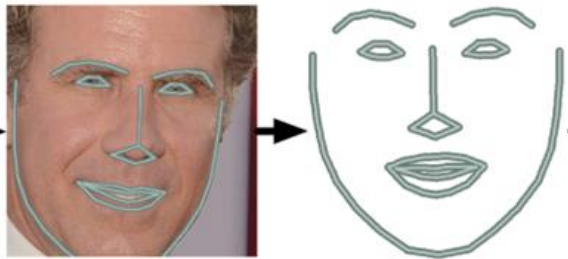


H-10. Dùng imglab để đánh dấu mốc trên khuôn mặt.
Ứng dụng này của Github có địa chỉ [imglab.ml](https://github.com/sergey-tyurin/imglab)

Sau đó xử lý tịnh tiến, co giãn và xoay để khuôn mặt để mắt mũi môi ở cùng một vị trí trong ảnh. Việc này giúp nhận diện ảnh tốt hơn.

3) Đo lường khuôn mặt bằng Convolution Neural Network (Quân, 2018)

Sau khi xử lý vấn đề xác định 68 điểm mốc trên khuôn mặt và dùng các phép tịnh tiến, co giãn và xoay các điểm mốc theo hướng nhìn thẳng để dễ cho việc nhận dạng. Tiếp theo nhận diện khuôn mặt khác nhau và cách tiếp cận đơn giản nhất là trực tiếp so sánh khuôn mặt tìm được ở trên.



H 11. Xử lý tịnh tiến, co giãn và xoay

Sử dụng mạng Deep Convolution Neural Network huấn luyện nó tìm ra 128 đo lường cho mỗi khuôn mặt. Quá trình huấn luyện thực hiện bằng cách cho máy tính nhìn 3 ảnh cùng một thời điểm (ảnh 1 và 2 cùng một người, ảnh 3 của người khác). Sau đó thuật toán sẽ khởi tạo ngẫu nhiên 128 giá trị đo lường. Tiếp đó nó thay đổi trọng số từ từ để tìm ra ảnh 1 và 2 ngày càng giống nhau, và giá trị từ ảnh 2 ngày càng khác với ảnh 3.

Khi lặp lại quá trình này hàng triệu lần với hàng triệu bức ảnh của hàng ngàn người khác nhau CNN học được cách tạo ra 128 đo lường cho mỗi người.

Cả ba phương pháp trên đều được sử dụng rộng rãi và giúp ích cho nhiều ứng dụng khác nhau của hệ thống nhận diện khuôn mặt nói chung. Và định danh người nói riêng.

B. Công cụ sử dụng

- Lập trình trên trình Visual Studio Code sử dụng ngôn ngữ Python phiên bản 3.7.
- Sử dụng các thư viện hỗ trợ như numpy, OpenCV, Dlib, Pickle.
- Thiết kế UI bằng thư viện PyQt5 trên Python.

C. Trình bày các độ đo sử dụng để đánh giá

Để đánh giá tỉ lệ nhận diện được ta tiến hành tính tỉ lệ giữa 128 đo lường được lưu trong data và 128 đo lường từ realtime. Giả sử:

128 đo lường được lưu trong data là `face_desc`

128 đo lường được từ real time là `face_desc0`

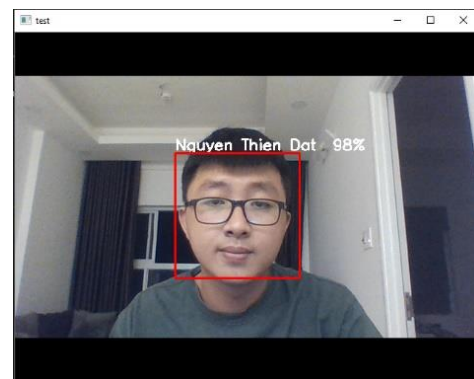
$$idx = \text{argmin}(\text{face_desc} - \text{face_desc0})$$

Nếu $idx < 0.3$ thì định danh được người và ngược lại sẽ không định danh được.

$$fid = \frac{\text{face_desc}}{\text{face_desc0}}, \text{ là độ chính xác của bài toán.}$$

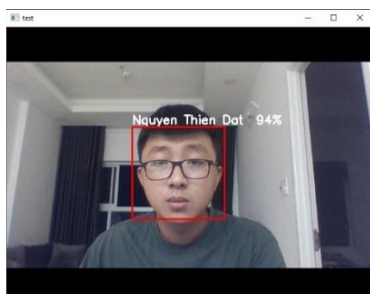
IV. Các thí nghiệm tinh chỉnh mô hình

A. Thử nghiệm với data chỉ có 1 cá nhân với 1 tấm ảnh.

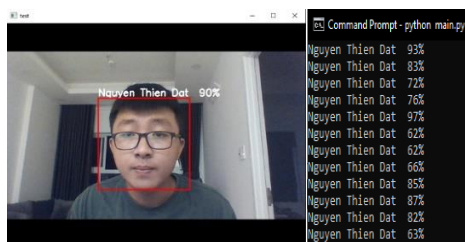


Nhận diện được 98% thuộc lớp Nguyễn Thiên Đạt

B. Thử nghiệm với 1 cá nhân với data 19 người mỗi người 1 tấm ảnh

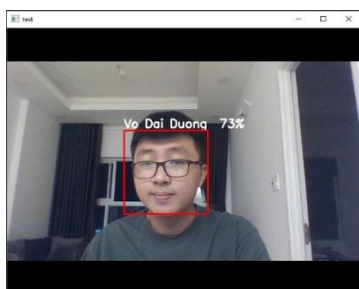


Nhận diện được 94% thuộc lớp Nguyễn Thiên Đạt

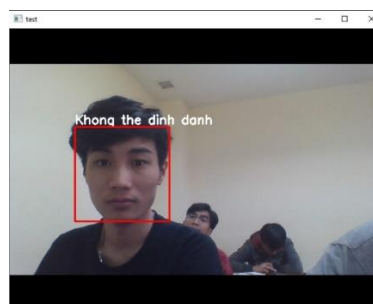


Quá trình nhận diện được với độ chính xác dao động từ 60-98%

D. Thử nghiệm với cá nhân không thuộc data

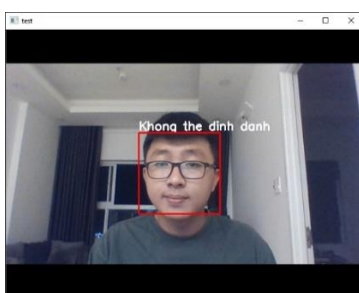


Nhận diện được 73% thuộc lớp Võ Đại Dương



Không tồn tại trong data thu thập nên không thể định danh được.

E. Giới thiệu phần mềm định danh khuôn mặt – Face Recognition



Không thể nhận diện được

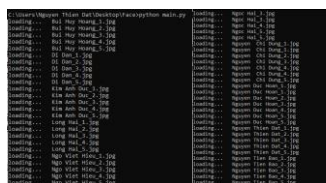


H-13. Face Recognition

- Background cơ bản.
- 1 layout.
- 2 button - training và camera mang chức năng như tên gọi.

Ta thấy với mỗi người một tấm ảnh sẽ ảnh hưởng đến độ chính xác của mô hình nhưng nhầm lẫn trong quá trình so sánh hay không thể định danh được. Nên ta sẽ sử dụng thêm mỗi người 5 tấm ảnh để tiến hành training và trích xuất khuôn mặt để tăng độ chính xác.

C. Thử nghiệm với 1 cá nhân với data mỗi người 5 tấm ảnh



H 12. Quá trình training và trích xuất khuôn mặt

V. Phân tích lỗi, hướng phát triển

A. Phân tích lỗi

- Còn hạn chế về UI
- Trong điều kiện ánh sáng quá chói hoặc quá tối sẽ không thể nhận diện được
- Góc nghiêng lớn hơn 10 độ sẽ không nhận diện được

B. Hướng phát triển

Ngày nay khoa học và công nghệ phát triển rất mạnh mẽ trong thời kì công nghiệp 4.0. Từ các ông lớn như Apple, Samsung đã đưa công nghệ nhận diện khuôn mặt để mở khoá điện thoại. Hay tại Việt Nam cũng đã có một số trường áp dụng để thực hiện điểm danh học sinh, sinh viên.

Vì vậy từ đề tài “*Phát hiện và định danh khuôn mặt*” làm nền tảng để ứng dụng cho việc mở khoá cửa văn phòng.

VI. Kết luận

Sau khi hiện thực đề tài, đã đúc kết được các kết quả như sau:

- Tìm hiểu được các phương pháp nhận dạng mặt người
- Hiểu được kỹ thuật xác định được 68 điểm mốc trên khuôn mặt và ứng dụng 68 điểm mốc đó vào quá trình nhận dạng khuôn mặt người
- Hiểu được kỹ thuật các nhà nghiên cứu mã hóa được 128 đo lường của khuôn mặt dựa trên 68 điểm mốc đã phát hiện.
- Có thể ứng dụng được các kỹ thuật để phát triển mục tiêu nghiên cứu cho bản thân
- Hiểu được kỹ thuật các nhà nghiên cứu mã hóa được 128 đo lường của khuôn mặt dựa trên 68 điểm mốc đã phát hiện.
- Đề tài còn nhiều thiếu sót do thời gian thực hiện ngắn nhưng nó cũng đạt được mục đích đề ra ban đầu.

TÀI LIỆU THAM KHẢO

- HaiHa. (2019, 3 16). *Viblo*. Retrieved from <https://viblo.asia/p/tim-hieu-ve-phuong-phap-nhan-dien-khuon-mat-cua-violas-john-ByEZkNVyKQ0>
- King, D. (2020). *GitHub*. Retrieved from <https://github.com/davisking/dlib-models>
- Masa. (2020). *Kaggle*. Retrieved from <https://www.kaggle.com/sajikim/shape-predictor-68-face-landmarks>
- Preda, G. (2020). *Kaggle*. Retrieved from [https://www.kaggle.com/gpreda/haar-](https://www.kaggle.com/gpreda/haar-cascades-for-face-)

detection?select=haarcascade_frontalface_default.xml

Quân, L. N. (2018, 4 3). Ứng dụng nhận dạng khuôn mặt trong thiết bị đóng cửa văn phòng. *Trường đại học Lạc Hồng*, (p. 19).