A PROJECT REPORT
ON

FACE RECOGNITION

BY

Trần Hoàng Nam
Thái Thị Thanh Linh
Lê Đức Trung

Internship - Sunshine Tech
Tp. Hồ Chí Minh, 7/2019

1 Giới thiệu chung

Project của nhóm bao gồm nghiên cứu, đánh giá và thực nghiệm các phương thức nhận diện khuôn mặt người đồng thời đưa ra phương án cải thiện model, tăng dộ chính xác, tốc độ xử lí và phù hợp với nhu cầu của người dùng

2 Xác định khuôn mặt

2.1 Phương pháp Haarcascade

Haarcascade được đề xuất bởi Paul Viola và Michael Jones năm 2001 sử dựng phương pháp chiết xuất đặc trưng Haarlike feature, chúng ta chuyển ảnh về dạng grayscale và phân tích độ intensity của từng pixel để xác định các Haarlike features

2.2 Phương pháp Dlib HOG

OG là viết tắt của Histogram of Oriented Gradient - một loại "feature descriptor". Mục đích của "feature descriptor" là trừu tượng hóa đối tượng bằng cách trích xuất ra những đặc trưng của đối tượng đó và bỏ đi những thông tin không hữu ích. Vì vậy, HOG được sử dụng chủ yếu để mô tả hình dạng và sự xuất hiện của một đối tượng trong ảnh.
Bản chất của phương pháp HOG là sử dụng thông tin về sự phân bố của các cường độ gradient (intensity gradient) hoặc của hướng biên (edge directins) để mô tả các đối tượng cục bộ trong ảnh. Các toán tử HOG được cài đặt bằng cách chia nhỏ một bức ảnh thành các vùng con, được gọi là "tế bào" (cells) và với mỗi cell, ta sẽ tính toán một histogram về các hướng của gradients cho các điểm nằm trong cell. Ghép các histogram lại với nhau ta sẽ có một biểu diễn cho bức ảnh ban đầu. Để tăng cường hiệu năng nhận dạng, các histogram cục bộ có thể được chuẩn hóa về độ tương phản bằng cách tính một ngưỡng cường độ trong một vùng lớn hơn cell, gọi là các khối (blocks) và sử dụng giá trị ngưỡng đó để chuẩn hóa tất cả các cell trong khối. Kết quả sau bước chuẩn hóa sẽ là một vector đặc trưng có tính bất biến cao hơn đối với các thay đổi về điều kiện ánh sáng.

2.3 Phương pháp Dlib CNN

Gần đây Dlib đã cung cấp thêm các hàm xác định khuôn mặt dựa trên mạng CNN(mạng lưới nerual) sử dụng nhiều layers.

2.4 Thực nghiệm

Nhóm tiến hành đánh giá xác định khuôn mặt thông qua hình ảnh qua các phương pháp với tập dữ liệu đầu vào bao gồm 50 ảnh chụp chính diện, 50 ảnh chụp nghiêng và 10 ảnh trong điều kiện không đủ ánh sáng.

Phương pháp	Số lượng ảnh xác định được			Tỉ lệ	Thời gian
	Ảnh chính diện	Ảnh nghiêng	Ảnh thiếu ánh sáng		
Haarcascade	28	2	0	27.3%	1,74s
Dlib HOG	50	32	4	78,2%	56,1s
Dlib CNN	50	50	8	98.2%	5,6s

2.5 Đánh giá

Dựa vào bảng thực nghiệm trên có thể đánh giá sơ bộ các phương pháp:

- Phương pháp Haarcascade: ưu điểm phương pháp này là thời gian xác định nhanh tuy nhiên khả năng xác định được khuôn mặt thấp nhất. Hầu như chỉ xác định được nếu đứng trực diện và không thể xác định khi điều kiện ánh sáng không tốt.
- Phương pháp Dlib HOG: xác định được tốt khi đứng trực diện và ổn ở góc nghiêng và điều kiện ánh sáng thấp. Nhược điểm tốn nhiều thời gian để xác định.
- Phương pháp Dlib CNN: là phương pháp xác định tốt nhất trong ba phương pháp trên, ổn định đều về thời gian và độ chính xác.

3 Nhận diện khuôn mặt

3.1 Thuật toán Eigenface

3.1.1 Hoạt động

Eigenface lấy ý tưởng đằng sau từ PCA, PCA là một phương pháp giảm chiều dữ liệu, khi mà dữ liệu có chiều lớn mà chúng ta chỉ có thể visualize ở chiều nhỏ hơn 3 thì PCA sẽ là một phương pháp giúp ta đữa data về một không gian mới(ta gọi là PCA space) mà vẫn cố giữ lại được thông tin nhiều nhất có thể trên data.

3.1.2 Thực nghiệm

Thuật toán có độ chính xác là 62% . Tương đối thấp vì PCA là 1 feature extraction dạng shadow learning nên feature chỉ làm việc tốt đối với những image có sự khác biệt lớn về structer and texture như chó mèo.. Còn face thì khó hơn ta có thể dùng các kỹ thuật feature của deep learning để training.

3.1.3 Đánh giá

Đơn giản và nhanh tuy nhiên khả năng nhận diện giảm dưới những điều kiện về kiểu dáng và ánh sáng khác nhau, database càng lơn thì khả năng nhận diện càng giảm, giải thuật yêu cầu về sự tương đồng về background

3.2 Thuật toán Fisherface

3.2.1 Hoạt động

Fisherface là một phương pháp làm việc với tập training mà mỗi đối tượng có nhiều ảnh mặt ở các điều kiện khác nhau). Sau khi thực hiện chiếu tập reference vào không gian con, hệ thống lưu lại kết quả là một ma trận với mỗi cột của ma trận là một vector tương ứng với ảnh (định danh đã biết) để thực hiện nhận dạng (hay phân lớp). Nhận dạng (hay phân lớp) được thực hiện với tập các ảnh probe, sau khi tiền xử lý xong, mỗi ảnh sẽ được áp dụng phương pháp trích chọn đặc điểm (như với các ảnh thuộc tập training và reference) và được chiếu vào không gian con. Tiếp đến việc phân lớp sẽ dựa trên phương pháp k-NN, định danh của một ảnh cần xác định sẽ được gán là định danh của ảnh có khoảng cách (distance) gần với nó nhất. Ở đây cần lưu ý là mỗi ảnh là một vector nên có thể dùng khái niệm hàm khoảng cách giữa hai vector để đo sự khác biệt giữa các ảnh.

3.2.2 Thực nghiệm

So sánh cách tiếp cận Eigenface và Fisherface theo quy mô của dữ liệu đào tạo và theo tư thế hình ảnh. Thử nghiệm cả hai thuật toán trên 20 hình ảnh cho số lượng tư thế khác nhau trong dữ liệu đào tạo. Fisherface tốt hơn so với phương pháp eigenface khi số lượng tư thế ít hơn Nhưng khi số lượng tăng lên, tỷ lệ nhận biết trong cả hai trường hợp là gần như giống nhau.

3.2.3 Đánh giá

Khả năng nhận diện ổn với điều kiện ánh sáng, trạng thái khuôn mặt khác nhau, giải thuật sử nhiều bộ nhớ (mỗi người phải có ít nhất 4-5 tấm hình mới có thể nhận dạng được và 10 tấm để nhận dạng ổn)

3.3 Thuật toán LBPH

3.3.1 Hoạt động

Về cơ bản, trong LBP, mỗi pixel sẽ được biểu diễn bởi 1 chuỗi nhị phân, giá trị thập phân của chuỗi chính là giá trị của pixel đó. Chuỗi nhị phân được sinh ra qua việc so sánh độ xám của các pixel xung quanh với pixel trung tâm. Nếu lớn hơn hoặc bằng thì là 1, nhỏ hơn thì là 0. giá trị thập phân của chuỗi nhị phân này chính là giá trị của pixel trung tâm trong sự biểu diễn bởi toán tử LBP. LBP được sử dụng nhiều bởi chi phí tính toán thấp và khả năng phân tách bới nó chỉ phát hiện những mẫu vân cục bộ như các điểm, các điểm cuối đường thẳng, biên cạnh và góc. Ta sẽ sử dụng các vùng chồng lấp và toán tử LBP trong một lân cận 4 - LBP được tính toán trong các vùng chồng lấp này. Thêm vào đó, tăng cường sự mô tả đặc trưng (vân) của toàn bộ khuôn mặt bằng cách tính toán histogram LBP toàn cục trên toàn bộ ảnh khuôn mặt. Trong cách biểu diễn này, đặc trưng vân của toàn bộ ảnh khuôn mặt được mã hóa bởi LBP trong khi đặc trưng hình dáng của khuôn mặt được phục hồi bởi sự kết hợp các histogram LBP cục bộ.

3.3.2 Thực nghiệm

LBPH phân tích từng khuôn mặt trong tập huấn luyện một cách riêng biệt và độc lập.
Trong LBPH, mỗi hình ảnh được phân tích độc lập, trong khi phương thức eigenfaces xem xét toàn bộ tập dữ liệu. Phương thức LBPH có phần đơn giản hơn, chúng ta phân tích từng hình ảnh trong bộ dữ liệu cục bộ; và khi một hình ảnh chưa biết mới được cung cấp, chúng ta thực hiện phân tích tương tự trên nó và so sánh kết quả với từng hình ảnh trong bộ dữ liệu. Cách chúng ta phân tích hình ảnh là bằng cách mô tả các mẫu cục bộ ở mỗi vị trí trong hình ảnh.

3.3.3 Đánh giá

Khả năng nhận diện hiệu quả trong điều kiện môi trường được kiểm soát(ánh sáng,..), khi áp dụng tuật toán LBPH không nhất thiết phải resize các bức ảnh về cùng 1 kích thước đồng nhất, khi cần update bộ nhận diện chỉ cần thêm dữ liệu mới vào dataset mà không cần phải re-train

4 Nhận diện khuôn mặt thật/giả

4.1 Hướng nghiên cứu

Sử dụng thuật toán CNN-Convolution Neural Network, một cấu trúc phổ biến trong Deep Learning. Bài toán nhận biết khuôn mặt thật /giả sử dụng cấu trúc CNN để phân loại ảnh. Trong tập dữ liệu khuôn mặt thu thập được sẽ gồm 2 classes ('fake', 'real') để dùng cho bộ phân loại.
Bước đầu tiến hành tạo ra 1 model để giải quyết bài toán. Model được tạo ra chỉ với vài bộ lọc đơn giản. Mỗi bộ lọc được thêm vào các lớp Convolution, Max Pooling. Cuối cùng là thêm vào các lớp Fully Connected.
Sau khi tạo xong model, ta tiến hành training trên tập dữ liệu. Trong tập dữ liệu, mỗi hình ảnh khuôn mặt đều có 1 nhãn tương ứng được lưu trữ trong 1 danh sách **"labels"** (các nhãn này đều được mã hóa về dạng one-hot). Dữ liệu được phân vùng với 75% dùng cho training và 25% dùng cho testing.
Ngoài ra, trong đoạn chương trình để train model có sử dụng 1 kỹ thuật gọi là *Data Augmentation* để tạo ra nhiều dữ liệu training từ dữ liệu ban đầu. Điều này giúp tăng độ hiệu quả của bộ phân lớp vì dữ liệu training càng nhiều thì sẽ có bộ phân lớp tốt hơn.

4.2 Thực nghiệm

Kết quả đào tạo dựa trên tập dữ liệu gồm 900 hình ảnh cho khuôn mặt thật và 230 ảnh cho khuôn mặt giả.
Sau khi tiến hành train đến Epoch thứ 100, thu được kết quả như sau:

```
106/106 [==============================] - 4s 37ms/step - loss: 0.0460 - acc: 0.9823 - val_loss: 8.8501e-04 - val_acc: 1.0000
[INFO] evaluating network...
              precision    recall  f1-score   support

        fake       1.00      1.00      1.00        58
        real       1.00      1.00      1.00       225

    accuracy                           1.00       283
   macro avg       1.00      1.00      1.00       283
weighted avg       1.00      1.00      1.00       283
```

Kết quả training cho thấy độ chính xác của model trên tập training là 98,23% và trên tập testing là 100%.

Sau khi quá trình training kết thúc, 1 file chứa kết quả được xuất ra có tên là **liveness.model**. Chỉ cần gọi lại file model này trong chương trinhg nhận diện khuôn mặt để sử dụng chức năng nhận diện khuôn mặt thật hay giả.

4.3 Đánh giá

Kết quả nhận diện sử dụng model trên nhận diện tốt đạt khoảng 90%. Cụ thể:

- Khuôn mặt thật (hình ảnh bắt được trực tiếp từ webcam): nhận diện tốt, tuy nhiên có vài khoảnh khắc nhận sai là *'fake'*. Vấn đề có thể do ảnh hưởng của ánh sáng, chất lượng webcam. Nhưng vấn đề lớn nhất có thể là do tập dữ liệu để đào tạo chưa đủ đa dạng.

- Ảnh 2D (chỉ mới thử trên ảnh thẻ): kết quả nhận diện tốt đạt 100% *'fake'*.

- Ảnh trên điện thoại đưa vào trước webcam: nhận diện tốt 100% *'fake'*.
- Video trên điện thoại đưa vào trước webcam: nhận diện tốt đạt khoảng 90%. Cụ thể, đối với những video có chất lượng hình ảnh tốt, tổng thể màu da trắng, sáng cho kết quả nhận diện *'fake'* la 100% . Còn đối với những video có chất lượng kém hơn (hơi mờ), tổng thể màu da tối cho kết quả *'fake'* đạt 80%, *'real'* 20%.

5 Kết luận và hướng phát triển

5.1 Kết luận

Nhóm đã đặt được những yêu cầu cần thiết cho 1 phần mềm nhận diện gương mặt, chương trình được tối ưu theo độ chính xác

5.2 Hướng phát triển

Phân biệt thật giả đối với trường hợp mô hình mặt 3D, phân tích chiều sâu khuôn mặt.