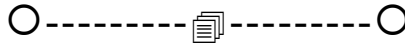


ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
KHOA HỆ THỐNG THÔNG TIN



MÔN HỌC: NHẬP MÔN THỊ GIÁC MÁY TÍNH
**BÁO CÁO ĐỒ ÁN: PHÁT HIỆN KHUÔN MẶT
TRONG ẢNH**

Giảng viên hướng dẫn : Thầy Mai Tiến Dũng

Lớp : CS406.O11.KHCL

Sinh viên thực hiện : Phạm Trần Xuân Khôi – 21521014

Thành phố Hồ Chí Minh, Tháng 01 năm 2024

[illegible]

Ký tên

MỤC LỤC

MỤC LỤC.....	3
Phần 1: GIỚI THIỆU BÀI TOÁN.....	4
I. Giới thiệu bài toán.....	4
II. Cấu trúc đồ án.....	4
Phần 2: PHƯƠNG PHÁP.....	5
I. Sliding Window.....	5
II. Trích xuất đặc trưng bằng HOG/LBP.....	5
III. Support Vector Machine (SVM).....	6
IV. Non-Maximum Suppression.....	6
Phần 3: CHUẨN BỊ DỮ LIỆU VÀ THỰC NGHIỆM.....	7
I. Bộ dữ liệu.....	7
II. Xử lý dữ liệu.....	7
III. Support Vector Machine (SVM).....	7
IV. Đánh giá mô hình.....	7
V. Sliding window và Non-Maximum Suppression.....	8
VI. Kết quả.....	8
Phần 4: KẾT LUẬN.....	9

Phần 1: GIỚI THIỆU BÀI TOÁN

I. Giới thiệu bài toán

Bài toán: nhận diện khuôn mặt có trong ảnh.

Input: Một hình ảnh một hoặc nhóm người có thấy được mặt.

Output: Hình ảnh với các bounding box nhận diện vị trí của các khuôn mặt có trong ảnh.

Bài toán vẫn sẽ hoạt động với hình ảnh không có người (trả về ảnh không có bounding box).

Nhận diện khuôn mặt là một lĩnh vực của thị giác máy tính liên quan đến việc tự động xác định và định vị khuôn mặt con người trong hình ảnh hoặc video kỹ thuật số. Đây là công nghệ cơ bản làm nền tảng cho nhiều ứng dụng như nhận dạng khuôn mặt, theo dõi khuôn mặt và phân tích khuôn mặt.

Nhận diện khuôn mặt mang lại nhiều tiềm năng ứng dụng trong nhiều lĩnh vực khác nhau, từ bảo mật và an ninh đến trải nghiệm người dùng và dịch vụ công nghệ.

II. Cấu trúc đồ án

Đồ án này có được chia thành các phần:

- Giới thiệu bài toán
- Phương pháp giải quyết
- Chuẩn bị dữ liệu và thực nghiệm
- Kết luận

Phần 2: PHƯƠNG PHÁP

Tổng quan phương pháp sử dụng:

1. Haar Cascade để detect khuôn mặt.
2. Sau đó phân loại giới tính bằng hai cách:
 - Sử dụng CNN
 - Sử dụng SVM

I. Sliding Window

Sliding Window là một kỹ thuật được sử dụng để quét và xác định các đối tượng trong ảnh hoặc video. Phương pháp này thường được áp dụng trong các nhiệm vụ như nhận dạng đối tượng, phân loại, hay theo dõi.

Kỹ thuật này có thể gặp thách thức khi áp dụng cho các ảnh có kích thước lớn hoặc khi đối tượng có kích thước và vị trí biến đổi đáng kể.

II. Trích xuất đặc trưng bằng HOG, LBP

1) Đặc trưng HOG

HOG là viết tắt của "Histogram of Oriented Gradients," và nó là một phương pháp trích xuất đặc trưng được sử dụng chủ yếu trong thị giác máy tính và nhận dạng đối tượng.

Các khái niệm về HOG được nêu ra từ năm 1986 tuy nhiên cho đến năm 2005 HOG mới được sử dụng rộng rãi sau khi Navneet Dalal và Bill Triggs công bố những bổ sung về HOG. HOG được tính toán trên một lưới dày đặc các cell và chuẩn hóa sự tương phản giữa các block để nâng cao độ chính xác.

2) Đặc trưng LBP

Local Binary Pattern-LBP là phương pháp trích chọn kết cấu của ảnh thành vector đặc trưng gọi là đặc trưng LBP. Ý tưởng ban đầu của phương pháp này

được tác giả Ojala và các đồng nghiệp giới thiệu, ban đầu xử lý trên ảnh xám. Ví dụ với một điểm ảnh xét 8 điểm lân cận và sử dụng chính giá trị của điểm ảnh tại trung tâm để phân ngưỡng 8 giá trị lân cận. Kết quả tìm được một chuỗi bit nhị phân có chiều dài bằng 8 tương ứng với 8 điểm lân cận được định nghĩa theo thứ tự nhất định. Chuỗi bit sẽ được chuyển sang hệ thập phân và giá trị thập phân này thay thế giá trị ban đầu của điểm ảnh đang xét. Số điểm lân cận có thể được thay đổi bằng đại lượng bán kính.

Ưu điểm của phương pháp LBP là có chi phí tính toán thấp, ổn định khi cường độ thay đổi đơn điệu và dễ mở rộng lên không gian nhiều chiều như ảnh màu trong hệ RGB.

III. Support Vector Machine (SVM)

Phương pháp học máy được sử dụng rộng rãi trong các bài toán phân loại và hồi quy.

Hoạt động bằng cách tìm ra một siêu phẳng trong không gian đa chiều sao cho khoảng cách từ siêu phẳng này đến các điểm dữ liệu gần nhất trên mỗi lớp là lớn nhất.

SVM đã được chứng minh là một trong những thuật toán phân loại tốt nhất trong nhiều bài toán thực tế.

IV. Non-Maximum Suppression

Non-Maximum Suppression là một kỹ thuật được sử dụng để giảm lượng dữ liệu đầu ra của mô hình nhận dạng đối tượng, đặc biệt là trong các nhiệm vụ như phát hiện vật thể hoặc đối tượng trong ảnh.

Khi một mô hình phát hiện vật thể được áp dụng, nó có thể tạo ra nhiều đề xuất cho mỗi đối tượng, và các đề xuất này có thể chồng chéo hoặc gần nhau. NMS được thiết kế để giữ lại chỉ những đề xuất có độ tin cậy cao nhất và loại bỏ những đề xuất không cần thiết.

Phần 3: CHUẨN BỊ DỮ LIỆU VÀ THỰC NGHIỆM

I. Bộ dữ liệu

Dataset cho nhãn “face”: Flickr-Faces-HQ Dataset (FFHQ):

1. Bao gồm 70000 hình ảnh khuôn mặt.
2. Sử dụng 10000 ảnh.

Dataset cho nhãn “non-face”: Caltech-256:

- Bao gồm 30607 hình ảnh thuộc 257 lớp khác nhau.
- Không sử dụng ảnh thuộc các lớp ‘people’ và ‘faces-easy-101’, còn lại 29963 ảnh.

=> Tổng cộng 39963 ảnh

Chia tỉ lệ 80/20 cho tập train/test

II. Xử lý dữ liệu

Resize ảnh về 64x64x1.

Tính vector đặc trưng HOG và LBP cho từng ảnh, sử dụng hàm tính HOG và LBP của skimage (các siêu tham số được khởi tạo theo mặc định).

III. Support Vector Machine (SVM)

Sử dụng các siêu tham số như sau:

- kernel = ‘poly’, degree={‘1’, ‘2’, ‘3’, ‘4’}
- kernel = ‘rbf’, C={‘0.5’, ‘1’, ‘2’, ‘4’}

IV. Đánh giá mô hình:

Độ đo đánh giá mô hình: F1 – Score

$$F_1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

kernel=‘poly’, degree=	1	2	3	4
LBP	0.9821	0.9777	0.9799	0.9804
HOG	0.9939	0.9968	0.9980	0.9978

kernel='rbf, C=	0.5	1	2	4
LBP	0.9781	0.9809	0.9826	0.9833
HOG	0.9971	0.9978	0.9983	0.9980

V. Sliding window và Non-Maximum Suppression

1. Sliding Window:

Kích thước: chiều dài ảnh / 10.

Khoảng cách = kích thước / 5.

Phóng lớn Sliding Window 10 lần, mỗi lần 25%.

2. Non-Maximum Suppression:

Ngưỡng IOU = 0.1.

VI. Kết quả



Phần 4: Kết luận và hướng phát triển

Ưu điểm:

- Nhận diện được hầu hết các khuôn mặt (trên 90%).
- Phương pháp đơn giản.
- Không tốn quá nhiều thời gian và tài nguyên cho training.

Nhược điểm:

- Thời gian nhận diện lâu (10-30s).
- Không thể nhận diện tốt trên các ảnh có khuôn mặt chiếm diện tích rất nhỏ so với ảnh ($<10\%$ chiều dài ảnh).
- Đôi khi nhận diện các vật không phải khuôn mặt.
- Bounding box không hoàn toàn khớp với khuôn mặt.

Hướng phát triển:

- Sử dụng phương pháp khác có tốc độ nhanh hơn như Faster-RCNN.
- Tìm kiếm thêm dữ liệu về nhiều góc độ của khuôn mặt và dữ liệu ‘non-face’ giống với khuôn mặt (như tượng, hình vẽ,...).
- Cải thiện và tối ưu bounding box: bounding box nhỏ nhất có thể mà vẫn bao quanh hoàn toàn khuôn mặt.