

Đồ án cuối khóa môn học
Nhận Dạng Thị Giác Và Ứng Dụng
GV: Lê Đình Duy - Nguyễn Tấn Trần Minh Khang

Học viên: Huỳnh Nhật Lưu
Đề tài: Bài toán Person Re-Identification

- I. Link Github:
<https://github.com/luunhathuynh/NhanDangThongTinThiGiac/tree/master/Do%20An%20Cuoi%20Khoa>
- II. Tái nhận dạng nhân vật (Person Re-Identification): tìm cùng một nhân vật đã xuất hiện ở những camera khác nhau (góc quay các camera không giao nhau). Chương trình được thực hiện thông qua các bước: Tạo dataset từ các bộ dữ liệu có sẵn -> rút trích đặc trưng bằng nhiều loại đặc trưng khác nhau -> đánh giá kết quả thông qua các độ đo khác nhau

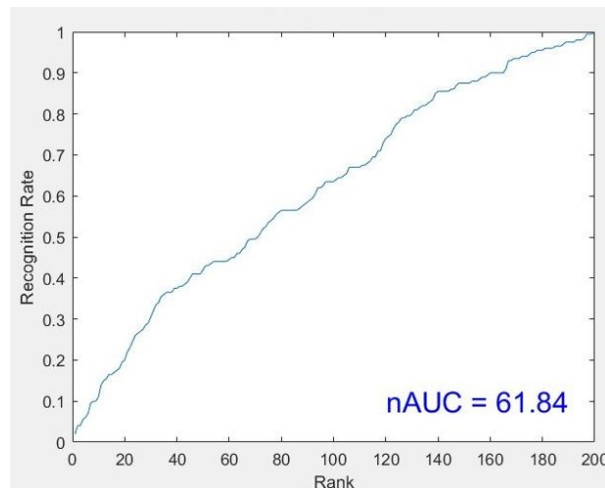
III. Nội dung

1. Các loại đặc trưng

- a. ELF : 29 loại đặc trưng, mỗi đặc trưng là 1 vector 16 chiều.
- b. HOG : vector đặc trưng 26244 chiều
- c. LBP : vector đặc trưng 59 chiều
- d. AlexNet (227x227) : lấy vector đặc trưng 4096 chiều ở lớp fc7
- e. VGG16 (224x224) : lấy vector đặc trưng 4096 chiều ở lớp fc7

2. Các độ đo

- a. ER : tổng các hạng (rank) đúng / tổng số mẫu.
- b. Rank 1 : tỷ lệ giữa số ảnh dự đoán đúng có hạng nhỏ hơn bằng 1 và tổng số ảnh.
- c. Rank 10 : tỷ lệ giữa số ảnh dự đoán đúng có hạng nhỏ hơn bằng 10 và tổng số ảnh.
- d. Rank 20 : tỷ lệ giữa số ảnh dự đoán đúng có hạng nhỏ hơn bằng 20 và tổng số ảnh.
- e. nAUC : được tính bằng diện tích dưới đường cong CMC



Biểu đồ mô tả đường cong CMC và độ đo nAUC

3. Các bộ dữ liệu

a. VIPeR

- Năm công bố: 2007
- Môi trường ghi nhận: ngoài trời
- Số người: 632
- Số ảnh: 632 (cam a) + 632 (cam b)
- Số camera: 2
- Số bounding box: 1264

b. PRID

- Năm công bố: 2011
- Môi trường ghi nhận: ngoài trời
- Số người: 934 trong đó chỉ có 200 người xuất hiện trong cả hai camera
- Số ảnh: 385 (cam a) + 749 (cam b)
- Số camera: 2
- Số bounding box: 24.541

c. GRID

- Năm công bố: 2009
- Môi trường ghi nhận: đường ngầm
- Số người: 250
- Số ảnh: 250 (cam a) + 1025 (cam b)
- Số camera: 8
- Số bounding box: 1275

d. PRID_450s

- Năm công bố: 2014
- Môi trường ghi nhận: ngoài trời
- Số người: 450
- Số ảnh: 1350 (cam a) + 1350 (cam b) (trong bài toán chỉ sử dụng 450 ảnh đầu từ cam a và 450 ảnh đầu từ cam b)
- Số camera: 2

4. Bảng đánh giá với từng dataset

	ER	R1	R10	R20	nAUC
ELF	142.76	7.59	18.35	27.37	77.48
HOG	271.59	0.32	3.64	6.17	57.1
LBP	291.9	0.95	3.8	6.01	0.54
AlexNet	162.92	3.32	12.34	18.04	74.3
VGG16	199.16	2.06	6.49	12.66	68.56

Dataset VIPeR

	ER	R1	R10	R20	nAUC
ELF	166.07	0.26	4.94	10.13	56.99
HOG	377.85	0	0.95	2.37	41.81
LBP	326.79	0	1.3	2.6	31.78
AlexNet	387.84	0.16	1.74	3.32	40.35
VGG16	387.36	0.32	1.11	3.01	40.51

Dataset PRID

	ER	R1	R10	R20	nAUC
ELF	123.30	0.8	8	12.80	50.88
HOG	508.22	0	0.47	1.74	30.72
LBP	438.72	0.4	1.6	2.4	19.59
AlexNet	515.03	0.16	0.95	2.22	30.74
VGG16	500.88	0	1.11	1.9	31.38

Dataset GRID

	ER	R1	R10	R20	nAUC
ELF	176.78	2.22	8.22	12.89	60.82
HOG	124.08	6.67	20	26.22	72.53
LBP	194.39	0.67	5.78	9.78	56.91
AlexNet	97.83	10.89	27.56	35.56	78.36
VGG16	125.59	4.67	19.56	27.78	72.2

Dataset PRID_450s

5. Nhận xét

- Qua bảng đánh giá, theo độ đo nAUC cho thấy:
 - Với các dataset VIPeR, PRID, GRID thì đặc trưng ELF cho độ chính xác cao nhất
 - Với dataset PRID_450s thì kiến trúc AlexNet cho độ chính xác cao nhất