TRƯỜNG ĐẠI HỌC SƯ PHẠM TP HỒ CHÍ MINH
TẠP CHÍ KHOA HỌC
KHOA HỌC TỰ NHIÊN VÀ CÔNG NGHỆ
ISSN: 1859-3100 Tập 15, Số 6 (2018): 97-106

HO CHI MINH CITY UNIVERSITY OF EDUCATION
JOURNAL OF SCIENCE
NATURAL SCIENCES AND TECHNOLOGY
Vol. 15, No. 6 (2018): 97-106

Email: tapchikhoahoc@hcmue.edu.vn; Website: http://tckh.hcmue.edu.vn

ĐÁNH GIÁ MỘT SỐ PHƯƠNG PHÁP BIỂU DIỄN ĐẶC TRƯNG CHO BÀI TOÁN TÁI NHẬN DẠNG NHÂN VẬT

Võ Duy Nguyên*, Nguyễn Thị Bảo Ngọc, Nguyễn Tấn Trần Minh Khang

Phòng Thí nghiệm Truyền thông Đa phương tiện - Trường Đại học Công nghệ Thông tin – ĐHQG TPHCM

Ngày nhận bài: 14-5-2018; ngày nhận bài sửa: 29-5-2018; ngày duyệt đăng: 19-6-2018

TÓM TẮT

Tái nhận dạng nhân vật là bài toán tìm kiếm các đối tượng đã di chuyển qua các camera khác nhau. Trong bài báo này, chúng tôi đánh giá thực nghiệm trên bộ dữ liệu lớn Airport, DukeMTMC4ReID được công bố gần đây bằng các phương pháp rút trích đặc trưng ELF, gBiCov, LOMO, WHOS. Kết quả cho thấy đặc trưng gBiCov có nAUC 54,42% (Airport), 40,61% (DukeMTMC4ReID) cao hơn các đặc trưng khác.

Từ khóa: tái nhận dạng nhân vật, hệ thống giám sát.

ABSTRACT

Empirical evaluation of feature representation methods for Person reidentification

Person re-identification is a practical task matching people moving across cameras. In this paper, we evaluated performance of various person re-identification approaches on recently published datasets Airport and DukeMTMC4ReId by feature extractors as ELF, gBiCov, LOMO, WHOS. The results show gBiCov achieved nAUC 54.42% (Airport), 40.61% (DukeMTMC4ReID) greater than the others.

Keywords: person re-identification, surveillance system.

1. Giới thiệu

Việc giám sát an ninh ở những nơi công cộng đang rất được chú trọng. Các camera giám sát được lắp đặt ở nhiều nơi như nhà ga, sân bay, trường học… Để vận hành các hệ thống giám sát này cần tốn nhiều chi phí về nhân lực và việc giám sát thủ công cũng không đảm bảo hiệu quả giám sát. Trong những năm gần đây, các hệ thống giám sát thông minh được xây dựng để nâng cao hiệu quả, giảm chi phí cũng như đáp ứng nhu cầu phát triển của các khu đô thị, thành phố thông minh. Bài toán Tái nhận dạng nhân vật (*person re-id*) là một trong những bài toán được ứng dụng trong việc giám sát an ninh.

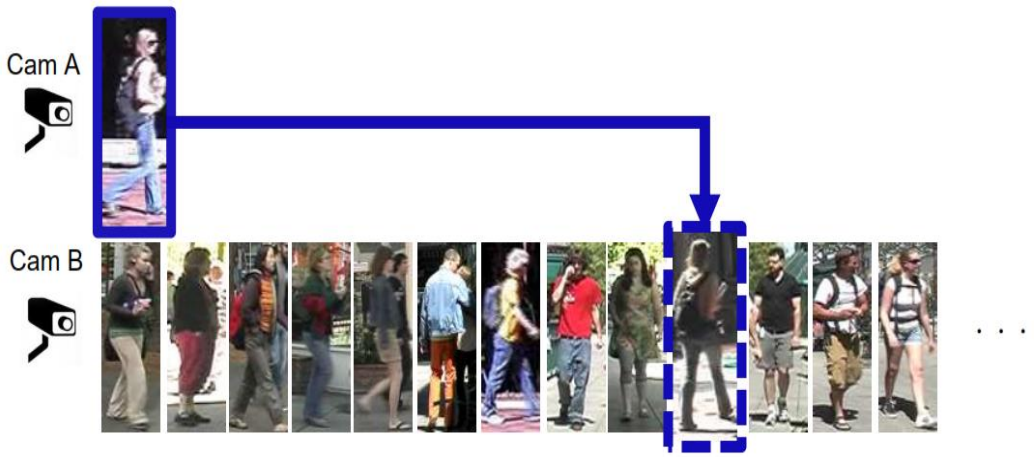
Tái nhận dạng nhân vật là bài toán có ảnh đầu vào là ảnh của một người thu được ở một camera, đầu ra là một danh sách những người được thu ở các camera khác, danh sách này được sắp xếp giảm dần theo mức độ tương đồng với ảnh đầu vào. Hình 1 minh họa bài toán tái nhận dạng nhân vật, tìm cùng một người xuất hiện ở hai camera khác nhau. Trong ví dụ, cô gái xuất hiện ở camera A ở góc quay ngang được tìm thấy ở camera B ở góc quay từ phía sau.

* Email: *nguyenvd@uit.edu.vn*

Bài toán có nhiều thách thức lớn do ảnh của nhân vật có nhiều biến thể trong những điều kiện khác nhau về ánh sáng, góc quay của camera, sự chồng lấp bởi các nhân vật hay các vật thể khác, sự thay đổi của nền (background) như trong nhà, ngoài trời hay giữa thời điểm ban ngày và ban đêm, thậm chí trong một số trường hợp thay đổi cả về trang phục của nhân vật.

Bài toán nhận được sự quan tâm của cộng đồng nghiên cứu thị giác máy tính trong hơn một thập kỉ qua [1, 2, 3]. Hai hướng nghiên cứu chính là biểu diễn đặc trưng (feature representation) và học độ đo khoảng cách (metric learning) giữa các đặc trưng. Bộ biểu diễn đặc trưng tốt sẽ "ổn định bền" trước những yếu tố làm đa dạng biến thể của nhân vật và giúp cho các phương pháp học độ đo khoảng cách giữa các hình ảnh biến thể của nhân vật đạt kết quả tốt hơn.

Cùng với sự phát triển của khoa học thế giới, nghiên cứu trong nước cũng có tiến triển, một số nghiên cứu sơ khởi nhằm nâng cao hiệu suất cho bài toán đã công bố [4, 5]. Tuy nhiên, ở Việt Nam chưa có một đánh giá nào trên các trên các bộ dữ liệu lớn mới được công bố trong những năm gần đây. Trong nghiên cứu này, chúng tôi sẽ trình bày tổng quan về các phương pháp biểu diễu đặc trưng và học độ đo khoảng cách để đánh giá trên những bộ dữ liệu mới bằng các độ đo tiêu chuẩn. Thông qua khảo sát này, chúng tôi cung cấp cái nhìn tổng quan hơn về bài toán tái nhận dạng nhân vật.



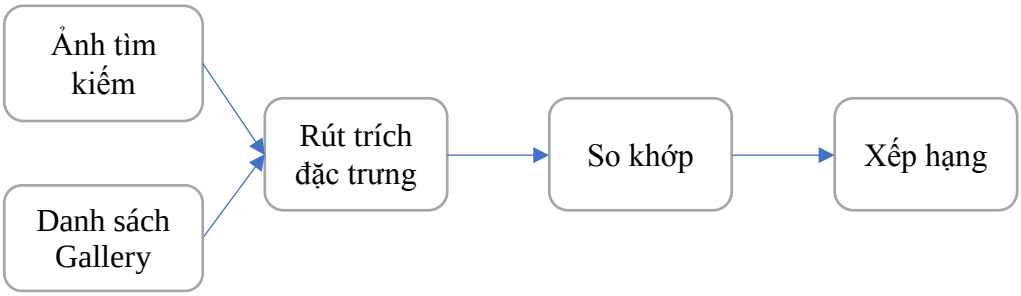
Hình 1. *Minh họa bài toán tái nhận dạng nhân vật [6]*

Phần còn lại của bài báo được tổ chức như sau: Phần 2 trình bày mô hình tái nhận dạng nhân vật, một số phương pháp rút trích đặc trưng và học độ đo. Phần 3 trình bày các bộ dữ liệu Airport, DukeMTMC4ReID, các độ tiêu chuẩn Rank i, nAUC và kết quả thực nghiệm. Cuối cùng, phần 4 trình bày kết luận.

2. Các nghiên cứu liên quan

2.1. Mô hình phổ biến của bài toán tái nhận dạng nhân vật

Trong phần này, chúng tôi trình bày tổng quan về các phương pháp rút trích đặc trưng của bài toán tái nhận dạng nhân vật. Tái nhận dạng nhân vật được nghiên cứu chủ yếu dựa trên ảnh đơn (single image). Bài toán được xem xét bao gồm dữ liệu 'gallery' có N ảnh tương ứng cho N người khác nhau (G_1, G_2,…, G_N) và dữ liệu ''Probe'' cũng có N ảnh, tương ứng với N người khác nhau (P_1, P_2,…, P_N), trong đó, ảnh G_1 và P_1 là cùng một người, tương tự cho các ảnh còn lại. Bài toán đặt ra là cho ảnh truy vấn q thuộc 'Probe' và tìm người giống ảnh q trong bộ dữ liệu 'gallery'. Các hướng giải quyết của bài toán chủ yếu xoay quanh hai vấn đề: một là **biểu diễn nhân vật** và hai là **so khớp các nhân vật.**



Hình 2. *Mô hình phổ biến của bài toán tái nhận dạng nhân vật*

Đặc trưng biểu diễn cho nhân vật được trích xuất từ ảnh thông qua các phương pháp rút trích đặc trưng. Một số phương pháp rút trích đặc trưng đã công bố trong các nghiên cứu trước đây: ELF, LDFV, gBiCov, IDE-CaffeNet, IDE-VGGNet, DenseColorSIFT, HistLBP, LOMO, GOG. Sau bước rút trích đặc trưng, chúng ta tiến hành so khớp (đặc trưng) các nhân vật. So khớp các đặc trưng để tính độ tương đồng của các nhân vật.

Một phương pháp truyền thống là tính khoảng cách Euclid (ℓ_2). Để tìm một người q trong tập dữ liệu 'gallery', chúng ta tính khoảng cách giữa (đặc trưng của) người q với tất cả người trong 'gallery', dựa vào kết quả khoảng cách, thu được danh sách sắp xếp giảm dần theo độ tương đồng. Những người đứng đầu danh sách gần giống với người q nhất. Thay vì sử dụng độ đo Euclid, một hướng tiếp cận khác là học có giám sát, phương pháp học độ đo khoảng cách (metric learning). Nhằm mục đích xác định những vectơ đặc trưng của cùng một người sẽ có khoảng cách gần hơn so với vectơ đặc trưng của những người khác.

2.2. Một số phương pháp rút trích đặc trưng

Đặc trưng thủ công được thiết kế bởi các chuyên gia để biểu diễn nhân vật dựa trên các đặc điểm của đối tượng. Như Ensemble of Localized Features (**ELF**) [7] được đề xuất bởi D. Gray and H. Tao vào năm 2008. ELF là đặc trưng kết hợp, sử dụng thông tin histogram màu của các kênh màu RGB, YcbCr và HS và các thông tin về kết cấu bề mặt

(texture) ảnh. Vectơ đặc trưng ELF dùng 29 đặc trưng gồm 8 kênh màu và 21 thông tin cấu trúc, mỗi đặc trưng là một vecto 16 chiều.

Gabor filters, Biologically Inspired Features and Covariance descriptors (**gBiCov**) là một phương pháp trích xuất đặc trưng mới dựa trên Gabor filters, Biologically Inspired Features (BIF) kết hợp với phương pháp Covariance descriptors. B.Ma và cộng sự đã công bố đặc trưng gBiCov vào năm 2014. Đặc trưng gBiCov thu được bằng cách tính toán và mã hóa sự khác biệt giữa đặc trưng sinh học BIF ở các tỉ lệ khác nhau. Khoảng cách giữa các nhân vật được tính toán hiệu quả bởi độ đo Euclidean.

Local Maximal Occurrence (**LOMO**) được đề xuất bởi S.Liao và cộng sự tại hội nghị CVPR 2015. LOMO sử dụng thuật toán đa tỉ lệ Retinex xử lí các đặc trưng LBP và biểu đồ màu HSV. LOMO phân tích sự xuất hiện theo chiều ngang của các đặc trưng cục bộ, và tối đa hóa sự xuất hiện tạo ra sự diễn tả rõ ràng hơn trước sự thay đổi của đối tượng qua các góc nhìn khác nhau.

Weighted Histogram of Overlapping Stripes (**WHOS**) là đặc trưng tập trung vào người (foreground) trong bức ảnh, dựa trên việc loại bỏ nền (background) bằng phương pháp Epanechnikov Kernel. WHOS lấy được nhiều đặc trưng về người ở trong ảnh, sau đó lấy histogram của ảnh và nối với đặc trưng HOG của ảnh đã loại bỏ nền.

2.3. Một số độ đo khoảng cách (Metric learning)

Một số phương pháp metric learning như: KISSME, MFA, FDA, NSFT. **MFA** (Marginal Fisher Analysis) loại bỏ được các sai lầm do bộ dữ liệu không phải dạng Gaussian. MFA là phương pháp tham số đặc trưng cục bộ kết hợp với k hàng xóm, nên nó có khả năng tính toán phi tuyến tính. **FDA** (Fisher's discriminant analysis) được tổng quát hóa bằng LDA, một phương pháp được sử dụng trong thống kê, nhận dạng mẫu và máy học để tìm một sự kết hợp tuyến tính của các tính năng đặc trưng hoặc tách hai hoặc nhiều lớp của đối tượng hoặc các sự kiện. **NSFT** (Null Foley-Sammon Transform) giải quyết vấn đề Small Sample Size (SSS) bằng việc áp dụng không gian phân biệt (discriminative null space), trong đó các hình giống nhau thì phải nằm chung một điểm trên không gian đó, và các hình không giống nhau phải nằm ở điểm khác, bằng một phép chiếu.

Trong phần này, chúng tôi tìm hiểu phương pháp kissme và cài đặt phương pháp này. **KISSME** [8] là phương pháp phân biệt ảnh khác nhau dựa trên hàm phân phối Gaussian giúp học nhanh bộ dữ liệu, bên cạnh đó sử dụng ma trận hiệp phương sai (cov matrix) giúp tăng hiệu suất (performance). Để đánh giá các phương pháp rút trích đặc trưng ELF, WHOS cho bài toán tái nhận dạng nhân vật, chúng tôi chọn KISSME làm phương pháp tính độ đo khoảng cách.

3. Bộ dữ liệu

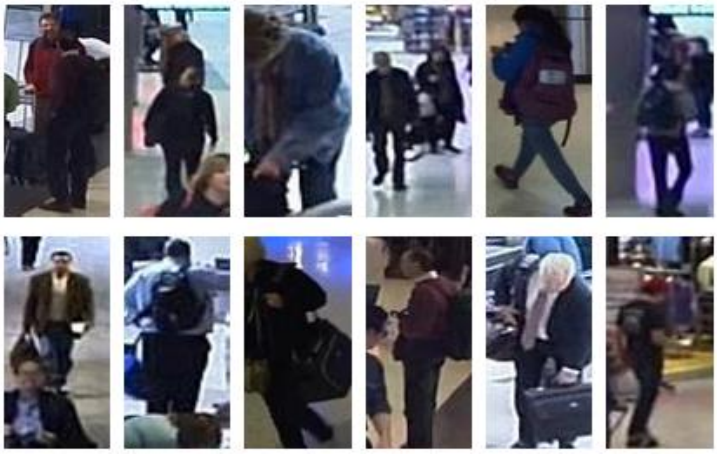
3.1. Bộ dữ liệu

Trong phần này, chúng tôi sẽ giới thiệu một số bộ dữ liệu được sử dụng để đánh giá thực nghiệm, bao gồm: Airport, DukeMTMC4ReID.

Bảng 1. *Các đặc tính 2 bộ dữ liệu: Airport, DukeMTMC4ReID*

STT	Tên bộ dữ liệu	Năm công bố	Số camera	Số người	Số ảnh	Môi trường lắp đặt
1	DukeMTMC4ReID	2017	8	1852	46261	Trường học
2	Airport	2018	6	1382	39902	Sân bay

Airport Bộ dữ liệu Airport [9] công bố năm 2018 được thu thập từ sáu camera giám sát lắp đặt ngoài trời ở một sân bay từ 8 giờ sáng đến 8 giờ tối. Ảnh người được nhận diện tự động với rất nhiều ảnh chỉ chứa một phần thân người (xem Hình 3). Tổng cộng, có 39.902 ảnh của 9651 người được thu thập. Trong đó có 1382 người xuất hiện trong ít nhất trong hai camera khác nhau. Airport là một bộ dữ liệu hứa hẹn với những đặc điểm giống như môi trường mở thực sự của một hệ thống giám sát thực sự.



Hình 3. *Minh họa bộ dữ liệu Airport*

DukeMTMC4ReID [10] là bộ dữ liệu mới nhất hiện nay được xây dựng dựa trên bộ dữ liệu DukeMTMC. Ảnh trong bộ dữ liệu DukeMTMC4ReID được thu thập từ một hệ thống bao gồm tám camera giám sát lắp đặt ở khuôn viên Trường Đại học Duke. DukeMTMC4ReID cung cấp 46.261 ảnh của 1852 người. Trong đó, 1413 người (22.515 ảnh) xuất hiện trong hơn một camera; 439 người còn lại (2195 ảnh) chỉ xuất hiện một trong tám camera. Ảnh trong bộ dữ liệu có kích thước giao động từ 72x34 pixel đến 515x188 pixel (xem Hình 4).

Hình 4. *Minh họa bộ dữ liệu DukeMTMC4ReID*

3.2. Độ đo

Bài toán tái nhận dạng nhân vật sử dụng độ đo chuẩn là Rank i, ngoài ra còn normalized Area under the CMC curve (nAUC). Các độ đo được xây dựng dựa trên **hạng đúng** (true rank). Tùy theo mỗi phương pháp biểu diễn đặc trưng và học độ đo khoảng cách sẽ có thứ tự xếp hạng khác nhau, phương pháp cho kết quả càng tốt thì người cần tìm trong ảnh query sẽ càng ở đầu danh sách. Thứ hạng của ảnh đúng trong danh sách xếp hạng là hạng đúng (true rank).

Rank i là khả năng dự đoán đúng trong i kết quả đầu tiên của Gallery. Rank i được tính bởi số lượng hạng đúng nhỏ hơn hay bằng i chia cho tổng số ảnh query. Rank i càng lớn thể hiện khả năng dự đoán đúng tại vị trí thứ i càng cao. Rank n luôn bằng 1, n là số ảnh query.

Đường cong Cumulative matching characteristic (CMC) vẽ tất cả các giá trị của Rank i. Trục hoành là dãy số nguyên cho biết thứ hạng i, trục tung là dãy số thực cho biết giá trị tương đương với mỗi thứ hạng i. Để so sánh kết quả giữa các đường cong CMC thì chúng ta sẽ dùng diện tích bên dưới đường cong CMC gọi là **normalized Area under the CMC curve (nAUC)**, giá trị lớn nhất của nAUC bằng 1.

3.3. Thực nghiệm

Trong phần này, chúng tôi trình bày kết quả thực nghiệm với bốn đặc trưng WHOS, LOMO, gBiCov và ELF. Mỗi đặc trưng có ưu và nhược điểm khác nhau. Để so sánh bốn đặc trưng, chúng tôi sử dụng cùng một metric learning KISSME và đánh giá trên hai bộ dữ liệu lớn, được công bố gần đây Airport, DukeMTMC4ReID. Tổ chức dữ liệu để huấn luyện và đánh giá cho bộ DukeMTMC4ReID được dựa vào tập tin đính kèm được tác giả bộ dữ liệu công bố. Dữ liệu Airport gồm 6 camera, chọn 1 camera làm ‘probe’ và chọn ngẫu nhiên cặp nhân vật từ 20 clip để làm dữ liệu huấn luyện, phần còn lại dùng cho đánh giá. Chúng tôi tiến hành thực nghiệm trên máy tính có cấu hình CPU intel(r) xeon(r) cpu

e5-2680, RAM 12GB, hệ điều hành Windows Server 2008 R2 Standard.

Bảng 2. *Mô tả thông tin các loại đặc trưng trên bộ dữ liệu Airport*

STT	Đặc trưng	Số chiều đặc trưng	Thời gian rút trích đặc trưng	Thời gian so khớp
1	WHOS	3,410	21 phút	17 phút
2	LOMO	4,044	1 tiếng 34 phút	21 phút
3	gBiCov	216	3 ngày 16 tiếng	15 giây
4	ELF	2,592	3 tiếng 48 phút	9 phút

Bảng 3. *Mô tả thông tin các loại đặc trưng trên bộ dữ liệu DukeMTMC4ReID*

STT	Đặc trưng	Số chiều đặc trưng	Thời gian rút trích đặc trưng	Thời gian so khớp
1	WHOS	3,410	25 phút	3 tiếng 31 phút
2	LOMO	4,044	1 tiếng 50 phút	3 tiếng 56 phút
3	gBiCov	216	4 ngày 14 tiếng	9 phút
4	ELF	2,592	3 tiếng 30 phút	1 tiếng 30 phút

Tổng quan về số chiều và thời gian chạy thực nghiệm các loại đặc trưng trên bộ dữ liệu Airport và DukeMTMC4ReID được tổng hợp trong bảng 3. Qua kết quả thực nghiệm Bảng 2 và Bảng 3, trên cả hai bộ dữ liệu Airport và DukeMTMC4ReID, mặc dù số chiều đặc trưng gBiCoV nhỏ nhất trong các đặc trưng nhưng thời gian rút trích đặc trưng gBiCoV nhiều nhất làm cho tổng thời gian rút trích và so khớp nhiều nhất.

Giá trị của Rank i tính theo %, nAUC là giá trị trong khoảng 0 và 1. Giá trị nAUC của các bộ dữ liệu có số lượng ảnh lớn như Airport và DukeMTMC4ReID sẽ tiến gần về giá trị 1. Trong thực nghiệm này, chúng tôi quan tâm đến 50 đầu tiên trong danh sách trả về và tính nAUC cho danh sách đó.

Đặc trưng WHOS cho kết quả Rank 1 cao nhất trên bộ dữ liệu Airport, tuy nhiên ở Rank 5,10 đặc trưng gBiCoV lại cho kết quả cao hơn WHOS và cũng cho giá trị nAUC cao nhất so với các đặc trưng được khảo sát. Trong khi đó trên bộ dữ liệu DukeMTMC4ReID thì đặc trưng gBiCoV cho kết quả Rank 1, 5, 10 cao nhất trong bốn đặc trưng (xem Bảng 4, 5).

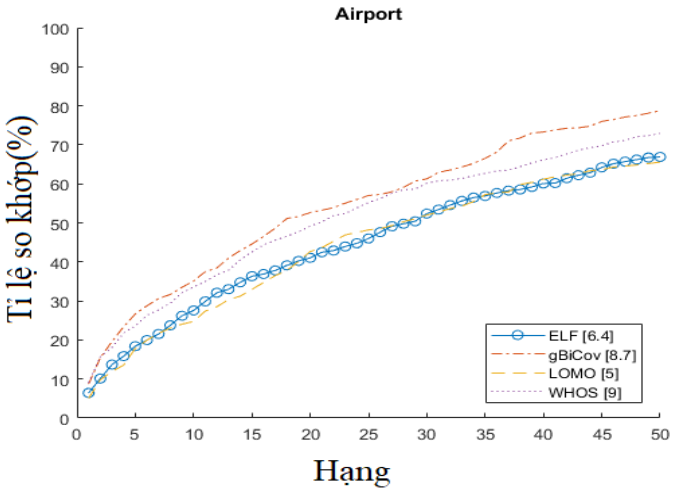
Kết quả từ Hình 5, 6 cho thấy đặc trưng gBiCoV cho kết quả tốt nhất trên trong 4 phương pháp biểu diễn đặc trưng trên cả hai bộ dữ liệu. Trên bộ Airport, ELF và LOMO hai đường cong tương đương nhau do có giá trị nAUC gần bằng nhau: ELF (44,53%) và LOMO (44,10%). Tuy nhiên, nếu dựa vào giá trị Rank 1 thì đặc trưng ELF (6,43) cho kết quả cao hơn LOMO (5,03).

Bảng 4. *Kết quả các độ đo trên bộ dữ liệu Airport*

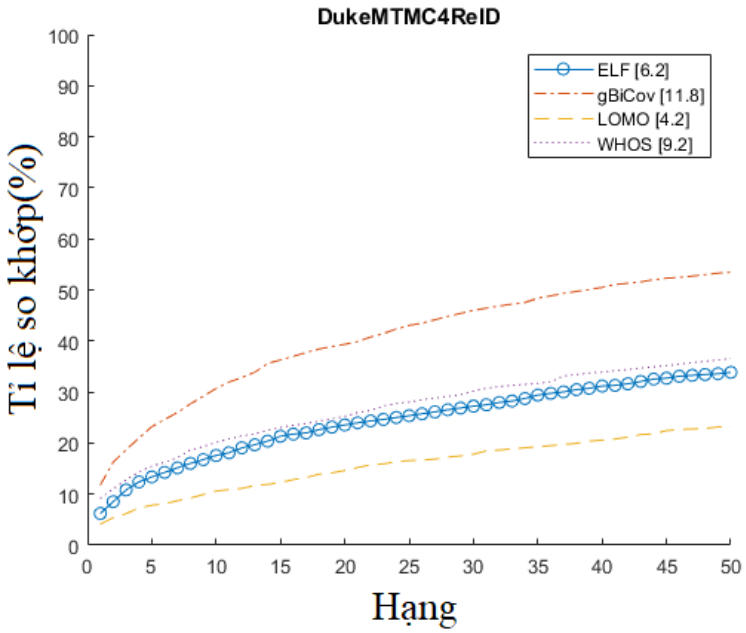
Đặc trưng	Rank 1	Rank 5	Rank 10	nAUC
WHOS	9,04	23,57	33,58	51,00
LOMO	5,03	17,83	24,73	44,10
gBiCov	8,69	26,65	35,11	54,42
ELF	6,43	18,4	27,56	44,53

Bảng 5. *Kết quả các độ đo trên bộ dữ liệu DukeMTMC4ReID*

Đặc trưng	Rank 1	Rank 5	Rank 10	nAUC
WHOS	9,17	15,47	20,13	26,86
LOMO	4,18	7,82	10,59	15,74
gBiCov	11,8	23,21	30,69	40,61
ELF	6,21	13,36	17,57	24,40



Hình 5. *Đường cong CMC cho các bộ dữ liệu Airport*



Hình 6. *Đường cong CMC cho các bộ dữ liệu DukeMTMC4ReID*

4. Kết luận và thảo luận

Trong nghiên cứu này, chúng tôi tìm hiểu và báo cáo thực nghiệm các phương pháp *biểu diễn nhân vật* và *so khớp các nhân vật* của bài toán Tái nhận dạng nhân vật trên hai bộ dữ liệu mới DukeMTMC4ReID và Airport. Các bộ đặc tả nhân vật dựa trên đặc trưng thủ công có ưu và nhược điểm khác nhau hướng tới từng đối tượng cụ thể, chịu ảnh hưởng lớn các đặc trưng của nhân vật và môi trường thu nhận ảnh. Trong nghiên cứu tiếp theo, chúng tôi sẽ tận dụng ưu điểm của từng đặc trưng để tạo một bộ đặc tả tốt hơn.

❖ ***Tuyên bố về quyền lợi:*** *Các tác giả xác nhận hoàn toàn không có xung đột về quyền lợi.*

❖ ***Lời cảm ơn:*** *Nghiên cứu này được thực hiện tại Phòng Thí nghiệm Truyền thông Đa phương tiện (MMLab) - Trường Đại học Công nghệ Thông tin - ĐHQG HCM (VNUHCM-UIT).*

TÀI LIỆU THAM KHẢO

[1] Y. Li, Z. Wu, S. Karanam, and R. Radke, "Real-world reidentification in an airport camera network," in ICDSC, 2014.

[2] D. Gray and H. Tao., "Viewpoint Invariant Pedestrian Recognition with an Ensemble of Localized Features," in European conference on computer vision (ECCV), 2008.

[3] M. Gou, X. Zhang, A. Rates-Borras, S. Asghari-Esfeden, M. Sznaier, and O. Camps, "Person re-identification in appearance impaired scenarios," in BMVC, 2016.

[4] N.-B. Nguyen, V.-H. Nguyen, T. N. Duc, D.-D. Le, and D. A. Duong, "AttRel: an approach

to person re-identification by exploiting attribute relationships," in in International Conference on Multimedia Modeling, pp. 50–60., 2015.

[5] N.-B. Nguyen, V.-H. Nguyen, T. D. Ngo, and K. M. T. T. Nguyen, "Person re-identification with mutual re-ranking," in Vietnam J. Comput. Sci., vol. 4, no. 4, pp.233–244,, 2017.

[6] Tetsu Matsukawa, Einoshin Suzuki, "Person Re-Identification Using CNN Features Learned from Combination of Attributes," in in Proceedings of International Conference and Pattern Recognition (ICPR2016), 2016.

[7] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in ECCV, 2008.

[8] M. Koestinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in CVPR, 2012.

[9] Srikrishna Karanam, Mengran Gou, Ziyan Wu, Angels Rates-Borras, Octavia Camps, Richard J Radke, "A Systematic Evaluation and Benchmark for Person Re-Identification: Features, Metrics, and Datasets," in IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018.

[10] Mengran Gou, Srikrishna Karanam, Wenqian Liu, Octavia Camps, Richard J. Radke, "DukeMTMC4ReID: A Large-Scale Multi-Camera Person Re-Identification Dataset," in CVPR, 2017.