

ĐỀ XUẤT GIẢI PHÁP BẮM ĐỐI TƯỢNG DỰA TRÊN NHẬN DẠNG SỬ DỤNG MÔ HÌNH TRÍ TUỆ NHÂN TẠO VÀ TRÍCH RÚT ĐẶC TRƯNG

Nguyễn Văn Tài, Lê Xuân Hưng
Cục Kỹ thuật nghiệp vụ - Bộ Công an

Tóm tắt- Trong bài báo này, chúng tôi trình bày một đề xuất về giải thuật bám các đối tượng người và phương tiện giao thông dựa trên nhận dạng đối tượng bởi mô hình trí tuệ nhân tạo, kết hợp phương pháp trích xuất đặc trưng HOG (Histogram of Oriented Gradients) của đối tượng cần được nhận dạng để tạo ra các vector đặc trưng HOG, các vector đặc trưng HOG được sử dụng làm đầu vào cho giải thuật bám đối tượng đề xuất. Tiếp đó, giải thuật bám đối tượng sử dụng các bộ lọc Kalman, kết hợp giải ma trận đường dẫn tối ưu Hungarian để duy trì việc bám các đối tượng của khung hình trước và khung hình hiện hành, và dự đoán các đối tượng được bám trong khung hình kế tiếp. Giải thuật đề xuất được tiến hành thử nghiệm trên tập dữ liệu MOTChallenge MOT16, được đánh giá so với 2 phương pháp gồm CLEAR MOT sử dụng các chỉ tiêu MOTA, MOTP, MT, ML, Flag; và Identity sử dụng các chỉ tiêu IDP, IDR, IDF1. Kết quả thử nghiệm đạt được tốt hơn so với các giải thuật OUTTrack_fm_p, MAT, Hugmot, XJTU đã công bố trước đó. Kết quả đạt được có độ chính xác bám đối tượng theo các tiêu chí MOTA, MOTP, MT, Flag ML lần lượt là 75.6%, 81.7%, 304, 31 và 1.8; và theo các tiêu chí IDP, IDR và IDF1 lần lượt đạt được là 81.5%, 72.7% và 76.8%.

Từ khóa- Bám đối tượng, nhận dạng, trích rút đặc trưng, HOG, MOTChallenge, CLEAR MOT, Identity.

I. GIỚI THIỆU

Trong khi các mô hình trí tuệ nhân tạo như YOLO, CAFFE, ONNX... được xây dựng tập trung cho mục tiêu nhận dạng và phân loại đối tượng, thì việc bám các đối tượng được nhận dạng thể hiện một khía cạnh phức tạp hơn của bài toán “thị giác máy tính”, bởi một thực tế rằng các đối tượng thay đổi rất đa dạng và phong phú về hành động, màu sắc, kích thước theo thời gian và không gian mà đối tượng xuất hiện. Vì vậy, tùy vào từng bài toán bám đối tượng cụ thể mà phải giải quyết một trong các yếu tố bất lợi

như đối tượng bị che khuất bởi một hoặc nhiều đối tượng khác, hình ảnh đối tượng bị biến dạng, mặt nền của đối tượng nhiễu xạ trộn, chiếu sáng vào đối tượng biến đổi, chuyển động nhanh, khung hình bị mờ do chuyển động, quay ngoài mặt phẳng, quay trong mặt phẳng, độ phân giải thấp hay đối tượng ngoài tầm nhìn. Để tăng cường khả năng khắc phục các khó khăn này, các giải thuật bám đối tượng hiện nay tập trung vào việc xử lý trực tiếp các đối tượng xuất hiện trong các khung hình để trích rút đặc trưng của đối tượng nhằm phân loại, chẩn đoán, nhóm, nhận dạng và phát hiện. Phương pháp trích xuất đặc trưng được sử dụng để thu được càng nhiều thông tin có thể về hình ảnh, kỹ thuật này được xem là một thách thức lớn hiện nay [1]. Có một số thuật toán được sử dụng để trích xuất các đặc trưng, gồm phân tích thành phần chính PCA [2], phân tích giá trị kỳ dị SVD [3] và phương pháp HOG [4-7]. Trong phương pháp PCA và SDV, đầu tiên chúng tính toán ma trận hiệp phương sai của tập dữ liệu, sau đó chọn một số cơ sở trực giao nhất định để xây dựng một không gian con có kích thước tương đối thấp hơn so với tập dữ liệu gốc, nhưng nó chỉ biểu thị thông tin toàn cục của đối tượng mà không xem xét thông tin cục bộ, do đó dẫn đến tỷ lệ nhận dạng thấp hơn. Phương pháp Histogram of Oriented Gradients (HOG) là một kỹ thuật mô tả tính năng phổ biến trong thị giác máy tính và xử lý hình ảnh. Nó phân tích sự phân bố các hướng cạnh trong một đối tượng để mô tả hình dạng và diện mạo của đối tượng đó. Phương pháp HOG bao gồm tính toán độ lớn và hướng của gradient cho mỗi pixel trong một hình ảnh và sau đó chia hình ảnh thành các ô nhỏ.

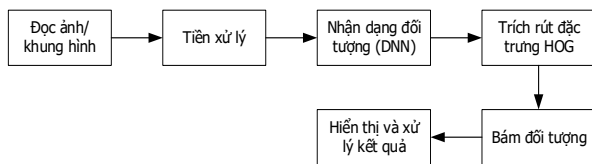
Trong bài báo này, chúng tôi đề xuất giải thuật bám đối tượng dựa trên nhận dạng sử dụng mô hình trí tuệ nhân tạo và trích rút đặc trưng đối tượng theo phương pháp HOG để tạo ra các vector đặc trưng HOG. Các vector đặc trưng này được sử dụng làm đầu vào cho giải thuật bám đối tượng đề xuất. Tiếp đó, giải thuật bám đối tượng sử dụng các bộ lọc Kalman, kết hợp giải ma trận đường dẫn tối ưu Hungarian để duy trì việc bám các đối tượng của khung hình trước và khung hình hiện tại, và dự đoán các đối tượng được bám trong khung hình kế tiếp.

Giải thuật của chúng tôi nhằm giải quyết những hạn chế của các phương pháp truyền thống như MIL [8],

Tác giả liên hệ: Nguyễn Văn Tài,
Email: tai2006vn@gmail.com
Đến tòa soạn: 11/9/2024, chỉnh sửa: 22/10/2024, chấp nhận đăng:
01/11/2024.

BOOSTING [9], MEDIANFLOW [10], TLD [11], KCF [12], GOTURN [13], và MOSSE [14], những phương pháp này thường gặp khó khăn trong việc xử lý các tình huống phức tạp như che khuất, biến đổi nhanh về hình dạng và kích thước, hoặc biến đổi ánh sáng. Bằng việc áp dụng trí tuệ nhân tạo để chính xác định vùng và phân loại đối tượng, giải thuật của chúng tôi cho phép một trích xuất đặc trưng đáng tin cậy hơn, từ đó nâng cao độ chính xác trong việc theo dõi đối tượng qua các khung hình. Yêu cầu cho mô hình AI trong phương pháp của chúng tôi là phải được huấn luyện một cách kỹ lưỡng, với khả năng thích ứng cao đối với các thay đổi về đối tượng và môi trường. Kết quả của giải thuật có thể đạt độ chính xác rất cao, nhất là khi mô hình được tinh chỉnh riêng biệt cho từng loại đối tượng cụ thể. Việc chọn lựa và phát triển các đặc trưng ảnh phù hợp, trong đó đặc trưng HOG đóng vai trò trung tâm, là yếu tố quan trọng quyết định đến hiệu quả của toàn bộ quá trình theo dõi, góp phần vào việc cải thiện độ nhạy và khả năng phân biệt của mô hình AI trong môi trường động.

Dựa vào cơ sở lý thuyết trên, sơ đồ giải pháp tổng thể như sau:



Hình 1. Giải pháp tổng thể bám đối tượng dựa trên nhận dạng sử dụng mô hình trí tuệ nhân tạo và trích rút đặc trưng đối tượng theo phương pháp HOG

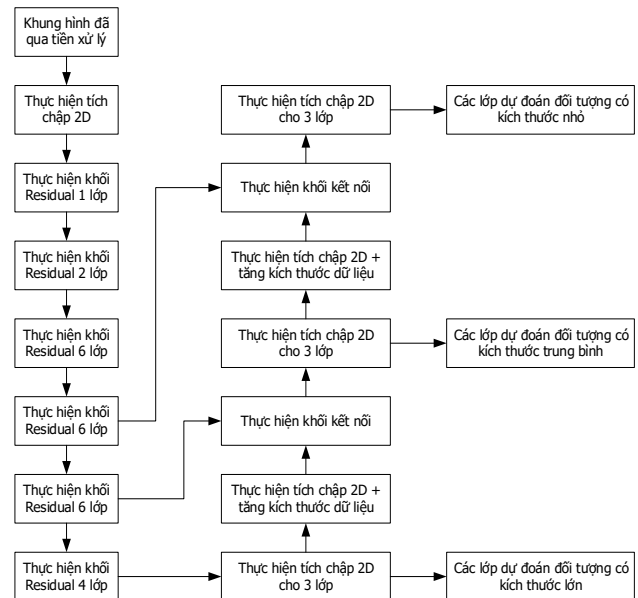
Hình 1 mô tả quy trình bắt đầu từ việc “Đọc ảnh/khung hình” nơi hình ảnh gốc được giữ nguyên không thay đổi, đảm bảo tính nguyên vẹn của dữ liệu. Tiếp theo, hình ảnh được “Tiền xử lý” để cải thiện chất lượng, bao gồm các thao tác như điều chỉnh độ sáng, tương phản và khử nhiễu. Sau đó, quy trình chuyển sang bước “Nhận dạng đối tượng (DNN)” sử dụng một mạng nơ-ron sâu để xác định và phân loại các đối tượng trong ảnh. Từ kết quả nhận dạng, “Trích rút đặc trưng HOG” được thực hiện, nơi các thông tin về hướng và độ lớn của gradient được sử dụng để tạo ra các vector đặc trưng giúp cải thiện độ chính xác của việc theo dõi hoặc nhận diện tiếp theo của đối tượng. Cuối cùng, hệ thống xử lý kết quả tổng hợp tất cả dữ liệu và “Bám đối tượng” để theo dõi sự di chuyển của đối tượng qua các khung hình tiếp theo. Quy trình này không chỉ cung cấp cách tiếp cận tổng thể trong việc nhận dạng và theo dõi đối tượng mà còn phù hợp với nhiều ứng dụng trong thực tiễn như giám sát an ninh, phân tích hành vi, và nhiều hơn nữa. Điểm nổi bật trong quy trình này có ba khối xử lý quan trọng nhất là: nhận dạng đối tượng (DNN), trích rút đặc trưng HOG, và bám đối tượng sẽ được trình bày chi tiết trong phần tiếp theo của bài báo: mục 2.1. Giải thuật thực hiện; mục 2.2. Thử nghiệm và đánh giá; mục 2.3. Kết luận.

II. NỘI DUNG CẦN GIẢI QUYẾT

A. Giải thuật thực hiện

1) Nhận dạng đối tượng

Khối “Nhận dạng đối tượng” trong biểu đồ sử dụng mô hình Deep Neural Network (DNN), cụ thể là một phiên bản cải tiến của mạng YOLO để phân tích và nhận dạng đối tượng trong hình ảnh [15]. Mô hình này hoạt động dựa trên một chuỗi các phân tích phức tạp, thực hiện qua nhiều lớp mạng nơ-ron, mỗi lớp có đặc điểm và chức năng khác nhau. Các cải tiến áp dụng cho mô hình bao gồm việc điều chỉnh kích thước ảnh đầu vào từ 416x416 xuống còn 208x208 điểm ảnh để tăng tốc độ xử lý. Ngoài ra, số lớp tích chập trong khối Residual được giảm từ 8 xuống còn 6 và số lớp tích chập 2D được giảm từ 5 xuống còn 3. Những thay đổi này nhằm mục đích tối ưu hóa hiệu suất tính toán mà vẫn duy trì độ chính xác cần thiết trong việc nhận dạng đối tượng. Mô hình đã được huấn luyện sử dụng một tập dữ liệu lớn đã được dán nhãn, cho phép nó học và nhận diện các đối tượng hiệu quả hơn. Sơ đồ giải thuật thực hiện các chức năng của khối DNN được thể hiện trong hình 2.



Hình 2. Sơ đồ giải thuật thực hiện các chức năng của DNN

2) Trích rút đặc trưng HOG

Đầu ra của khối “Nhận dạng đối tượng” (DNN) bao gồm một danh sách các đối tượng được phát hiện, cùng với thông tin chi tiết như tọa độ, kích thước vùng chiếm giữ, điểm số phân loại, và chỉ số loại đối tượng. Dựa vào các thông tin này, mỗi đối tượng được xác định sẽ được cắt ra thành một ảnh con từ ảnh đầu vào. Mỗi ảnh con này sau đó được sử dụng để trích xuất đặc trưng HOG (Histogram of Oriented Gradients). Quá trình này bao gồm việc tính toán các gradient (độ dốc) theo hướng và cường độ tại mỗi điểm ảnh, và sau đó tổng hợp các thông tin này thành một vector đặc trưng. Vector đặc trưng HOG này mang đầy đủ thông tin về hình dạng và cấu trúc của đối tượng trong không gian ảnh, hỗ trợ cho việc nhận diện và theo dõi đối tượng trong các bước xử lý tiếp theo.

Tính toán tốc độ biến thiên (gradient) của các điểm ảnh theo 2 phương ngang (x) và dọc (y): thực hiện việc lọc với 2 tập hệ số H_x (cho phương ngang x) và H_y (cho phương dọc y). $H_x = [-1, 0, 1]$, $H_y = [-1, 0, 1]$. Gọi A là mảng dữ liệu điểm ảnh 2 chiều có số cột $N_x = 64$, và số hàng $N_y = 128$. Tập giá trị tốc độ biến thiên theo x (cột), gọi là G_x . Tập giá trị tốc độ biến thiên theo y (hàng), gọi là G_y . G_x và G_y cũng là các mảng dữ liệu 2 chiều có cùng số hàng và số cột với A, và được tính theo công thức (1) và (2) [16].

$$G_x = H_x * A = [-1 \ 0 \ 1] * A$$

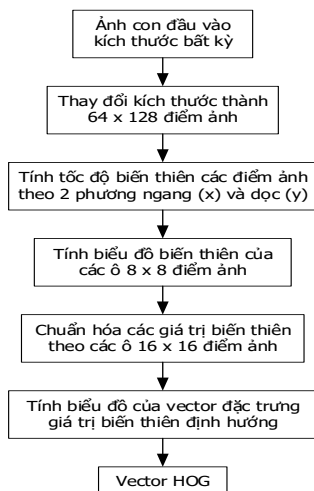
$$G_y = H_y * A = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} * A$$

Vì ảnh đầu vào là RGB với 3 kênh, nên các mảng A, G_x , G_y sẽ là các mảng dữ liệu với 3 kênh, mỗi kênh có 2 chiều. Sau khi G_x , G_y được tính toán, các tập dữ liệu về biên độ và góc hướng của tốc độ biến thiên sẽ được tính theo các công thức (3) và (4) [16].

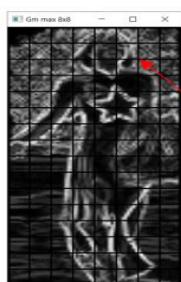
$$G_m = \sqrt{G_x * G_x + G_y * G_y}$$

$$G_\theta = \arctan\left(\frac{G_y}{G_x}\right)$$

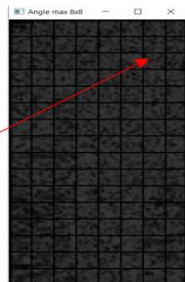
Vì mỗi ảnh có 3 kênh nên giá trị lớn nhất trong 3 giá trị của mỗi phần tử trong mảng biên độ G_m và góc G_θ sẽ được chọn để tạo ra mảng dữ liệu gm_max và $angle_max$ cho các bước xử lý sau.



Hình 3. Sơ đồ giải thuật thực hiện trích rút đặc trưng HOG



Hình 4. Ảnh gm_max



Hình 5. Ảnh $angle_max$

Tính biểu đồ biến thiên của các ô 8 x 8 điểm ảnh để xác định phân bố biên độ biến thiên theo khoảng góc: Các ảnh gm_max và $angle_max$ được chia thành các ô 8 x 8 điểm ảnh. Với kích thước 64 x 128 điểm ảnh, sẽ có 128 ô kích thước 8 x 8. Với mỗi ô 8 x 8, sẽ tính được một biểu đồ biến thiên biên độ theo 9 khoảng góc (9 bins) từ 0^0 tới 180^0 với giá trị bắt đầu của mỗi khoảng là: 0^0 , 20^0 , 40^0 , 60^0 , 80^0 , 100^0 , 120^0 , 140^0 , 160^0 .

Để không mất tính tổng quát, chọn ô 8 x 8 thứ 13, giá trị biên độ và góc tại mỗi điểm ảnh của ô này được hiển thị như Bảng 1, 2.

Bảng 1. Các giá trị biên độ của ô thứ 13

Gm cell: 13							
17	7	6	6	22	37	29	74
9	5	6	12	19	24	45	37
4	8	10	13	17	25	63	36
3	2	12	12	20	38	81	27
1	2	3	7	8	51	78	26
7	4	1	7	5	31	69	43
14	5	7	3	22	30	50	49
23	7	6	3	7	14	35	60

Bảng 2. Các giá trị góc của ô thứ 13

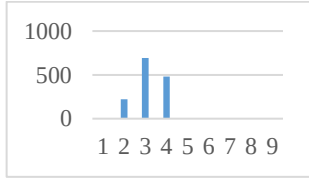
Angle cell: 13							
25	54	43	42	22	45	22	45
51	30	47	52	14	56	49	50
57	51	36	55	45	24	52	50
30	38	39	57	50	54	20	57
44	52	19	54	47	52	50	40
29	46	25	35	48	54	46	45
37	23	43	34	50	51	51	50
45	36	56	32	47	28	40	35

Gọi ô 8 x 8 bất kỳ của gm_max là gm_max_8x8 , ô 8 x 8 của $angle_max$ là $angle_max_8x8$. Nguyên tắc tạo biểu đồ biến thiên như hình 6 cho mỗi ô 8 x 8 như sau:

Tại vị trí xác định (i, j) của mảng $angle_max_8x8$, kiểm tra giá trị góc xem cách cận biên dưới và cận biên trên bao nhiêu độ, và tính tỉ lệ % khoảng cách đó so với 20^0 . Khoảng cách càng gần thì tỉ lệ càng lớn. Ví dụ, tại vị trí (0,0), giá trị góc là 25^0 , cách cận biên dưới 20^0 là 5^0 và cận biên trên 40^0 là 15^0 . Theo tỉ lệ %, thì 5^0 tương ứng với $100\% - 5/20 = 75\%$, và 15^0 tương ứng với $100\% - 15/20 = 25\%$.

Cũng tại vị trí xác định (i, j) của mảng gm_max_8x8 , lấy giá trị biên độ nhân với 2 tỷ lệ % của góc tại vị trí tương ứng, cho 2 giá trị tương đương với cận biên dưới và cận biên trên của góc. Ví dụ tại vị trí (0,0), giá trị biên độ là 17, tương ứng với cận biên dưới ta có giá trị 1 = $17 * 75.00\% = 12.75$, tương ứng với cận biên trên ta có giá trị 2 = $17 * 25\% = 4.25$

Áp dụng phương pháp trên cho 64 điểm ảnh của mỗi ô, sau đó tính tổng giá các giá trị biên độ tại mỗi khoảng góc. Cuối cùng, sẽ có một vector 9 phần tử cho một ô 8 x 8. Áp dụng tương tự cho toàn bộ 128 ô 8 x 8 của mảng gm_max và angle_max, ta sẽ có một tập hợp 128 vector 9 phần tử biểu diễn phân bố biên độ biến thiên theo khoảng góc. Dữ liệu thể hiện trên hình 6 là kết quả của tập hợp 128 vector 9 phần tử biểu diễn phân bố biên độ biến thiên theo khoảng góc, từ vector thứ 1 đến vector thứ 9 có các giá trị tương ứng lần lượt là 5.85, 221.50, 694.35, 482.30, 0.00, 0.00, 0.00, 0.00, 0.00.



Hình 6. Biểu diễn phân bố biên độ biến thiên theo khoảng góc

Chuẩn hóa các giá trị biến thiên theo các ô 16 x 16 điểm ảnh: Các đối tượng xuất hiện trên ảnh có thể biến đổi về kích thước và bị thay đổi về chiều sáng phụ thuộc vào vị trí đối tượng theo không gian và thời gian. Những thay đổi này cũng sẽ dẫn đến sự thay đổi về các giá trị biên độ biến thiên, dẫn đến những ảnh hưởng độ chính xác của việc trích rút đặc trưng. Để tối thiểu những ảnh hưởng này, các vector 9 phần tử này sẽ được chuẩn hóa theo kích thước các ô 16 x 16 điểm ảnh. Phương pháp thực hiện như sau:

Tập hợp 128 vector 9 phần tử được tổ chức thành một mảng dữ liệu gồm 16 hàng x 8 cột. Mỗi phần tử của mảng là một vector 9 phần tử. Tạo một cửa sổ gồm 4 ô liền kề nhau gồm 2 hàng x 2 cột. Từ cửa sổ này tạo ra một vector 36 phần tử bằng cách ghép 4 vector 9 phần tử của 4 ô liền kề. Mỗi phần tử của vector 36 phần tử được chuẩn hóa theo (5).

$$\|x_i\| = \frac{x_i}{\sqrt{\sum_{i=0}^{35} x_i * x_i}}$$

Tiếp tục dịch cửa sổ đi một cột theo chiều ngang tới khi số cột bằng 7, thì quay lại cột đầu với số hàng tăng lên 1 tới khi số hàng bằng 15. Mỗi lần dịch cửa sổ sẽ tạo ra một vector 36 phần tử được chuẩn hóa. Quá trình này lặp lại đến khi số cột bằng 7 và số hàng bằng 15. Tổng cộng sẽ tạo ra 7 x 15 = 105 vector 36 phần tử.

Ghép nối tiếp 105 vector 36 phần tử tạo được từ bước trên, sẽ tạo ra một vector 3780 phần tử. Vector 3780 phần tử này là biểu đồ HOG của ảnh đối tượng được chọn.

3) Bám đối tượng

Bám đối tượng được thực hiện dựa trên các nguyên tắc sau:

- Tập đối tượng được bám được khởi tạo là tập đối tượng được nhận dạng tại khung hình đầu tiên, mỗi đối tượng này được gán một số định danh duy nhất.
- Với các khung hình tiếp theo, nếu đối tượng bám ở khung hình trước được tìm thấy ở khung hình hiện tại, thì

số định danh của đối tượng tiếp tục được duy trì và đặc trưng của đối tượng được cập nhật với đặc trưng mới.

- Với các khung hình tiếp theo, nếu đối tượng bám ở khung hình trước không được tìm thấy ở một số khung hình liên tiếp tiếp theo, đối tượng bám được kết luận là đã bị mất. Đặc trưng của đối tượng bị mất được duy trì tiếp một số khung, sau đó sẽ được xóa khỏi danh sách.

- Với các khung hình tiếp theo, nếu có đối tượng nhận dạng mới và không tìm thấy trong danh sách các đối tượng đã bám, thì đối tượng được xem là đối tượng bám mới.

Thông tin được sử dụng cho việc bám đối tượng bao gồm: vị trí đối tượng và vùng chiếm giữ, vector đặc trưng HOG, và điểm phân loại đối tượng. Các thông tin này có được từ bước nhận dạng đối tượng DNN và trích rút đặc trưng HOG đã mô tả ở trên.

Việc so sánh và xác định quan hệ giữa các đối tượng giữa các khung hình liên tiếp nhau (được tìm thấy hoặc không được tìm thấy) dựa trên các cơ sở toán học sau:

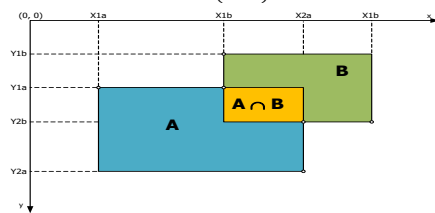
- Giá trị phép đo tương đồng cosine [17] (dcos) giữa các vector đặc trưng của các đối tượng. Giả thiết 2 vector A và B có kích thước là n, giá trị cosine giữa 2 vector này được tính theo (6, 7):

$$\cos(\theta) = S_c(A, B) = \frac{A \cdot B}{\|A\| \cdot \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \cdot \sqrt{\sum_{i=1}^n B_i^2}} \quad (6)$$

$$dcos = 1.0 - \cos(\theta) = 1.0 - S_c(A, B) \quad (7)$$

- Giá trị phép đo IoU [18] (Intersection over Union) cho thấy sự tương đồng (align) giữa vùng đối tượng hiện hành và vùng đối tượng đã bám được. Gọi vùng đối tượng hiện hành là A, vùng đối tượng đã bám được là B, giá trị IoU được tính theo (8).

$$(IoU) = \frac{d(A \cap B)}{d(A \cup B)} \quad (8)$$



Hình 7. Vùng chứa các đối tượng cần bám

Giả sử vùng đối tượng A và B như hình 7, trong đó:

- ✓ Đối tượng A có vùng chiếm giữ được xác định bởi tọa độ (X1a, Y1a) và (X2a, Y2a); đối tượng B có vùng chiếm giữ được xác định bởi tọa độ (X1b, Y1b) và (X2b, Y2b).
- ✓ Vùng giao của A và B được xác định bởi 2 tọa độ (X1b, Y1a) và (X2a, Y2b)
- ✓ Diện tích vùng đối tượng A: $d(A) = (X2a - X1a) \cdot (Y2a - Y1a)$
- ✓ Diện tích vùng đối tượng B: $d(B) = (X2b - X1b) \cdot (Y2b - Y1b)$
- ✓ Diện tích vùng giao của A và B: $d(A \cap B) = (X2a - X1b) \cdot (Y1b - Y1a)$

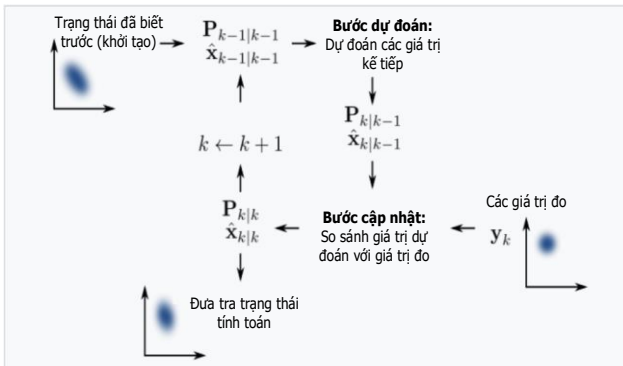
- ✓ Diện tích vùng hợp của A và B: $d(A \cup B) = d(A) + d(B) - d(A \cap B)$
- ✓ Và giá trị IoU sẽ là: $d(IoU) = \frac{d(A \cap B)}{d(A \cup B)}$

Giải bài toán tìm ánh xạ tối ưu giữa các đối tượng được nhận dạng của khung hiện hành và các đối tượng đã bám được ở khung trước đó, sử dụng giải thuật Hungarian [19]. Giả sử có một danh sách A gồm n các đối tượng được nhận dạng của khung hiện hành, và một danh sách B gồm m đối tượng đã bám được của khung thứ nhất. Lập một ma trận có kích thước m hàng và n cột là $D[m \times n]$, trong đó các phần tử là độ tương đồng cosine $dcos$ hoặc giá trị $d(IoU)$ giữa mỗi đối tượng được nhận dạng và mỗi đối tượng đã bám được. Biểu diễn toán học ma trận D như (9):

$$D = \begin{bmatrix} d_{A1B1} & d_{A1B2} & \dots & d_{A1Bn} \\ d_{A2B1} & d_{A2B2} & \dots & d_{A2Bn} \\ \vdots & \vdots & \ddots & \vdots \\ d_{AmB1} & d_{AmB2} & \dots & d_{AmBn} \end{bmatrix}$$

Bài toán được giải với điều kiện $m = n$, trong trường hợp $m \neq n$, hoặc một số hàng hoặc một số cột được thêm vào để đạt được $m = n$ với các phần tử có giá trị khoảng cách mặc định lớn nhất. Các giá trị d_{AiBj} hoặc sẽ là giá trị $dcos$ hoặc sẽ là giá trị $d(IoU)$.

Giải thuật Hungarian được sử dụng để tìm kiếm nhanh sao cho tổng các giá trị d_{AiBj} là nhỏ nhất với điều kiện chỉ có nhiều nhất 1 giá trị trong mỗi hàng và mỗi cột tham gia vào tổng này, nghĩa là không có từ 2 giá trị trở lên cùng hàng hoặc cùng cột.

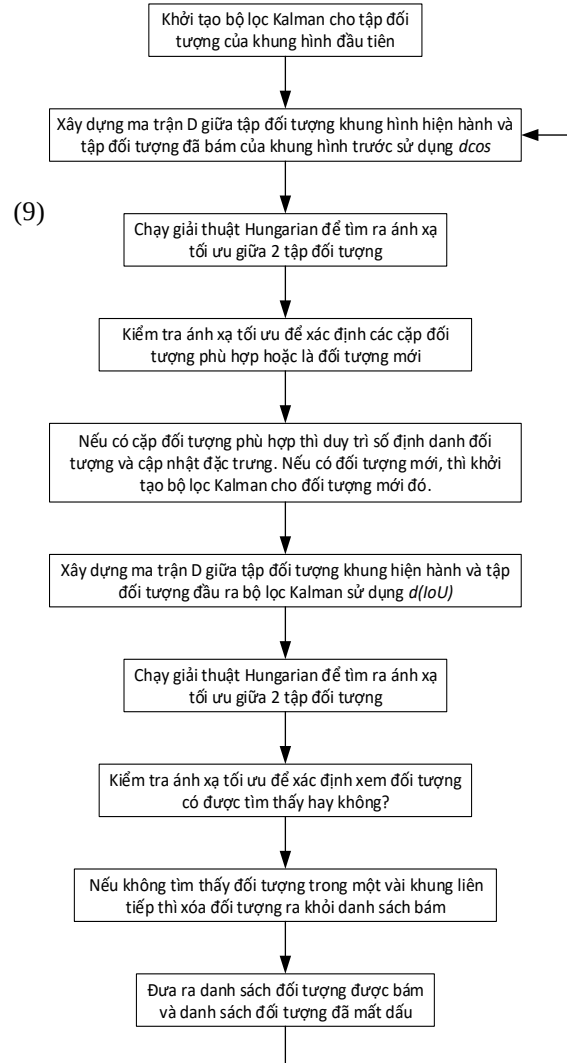


Hình 8. Sơ đồ tổng quát thực hiện bộ lọc Kalman

- ✓ Giá trị khoảng cách cosine $dcos$ được tính toán dựa trên vector HOG của mỗi đối tượng
- ✓ Giá trị $d(IoU)$ được tính toán dựa trên vị trí và vùng chiếm giữ của mỗi đối tượng.
- ✓ Điểm phân loại được sử dụng để làm ngưỡng xác định hoặc bỏ qua hoặc chấp nhận đối tượng khi lập ma trận D.

- Áp dụng bộ lọc Kalman [20-21] để dự đoán vị trí của đối tượng được bám trong khung hình kế tiếp được mô tả như hình 8. Bộ lọc Kalman sẽ sử dụng đặc trưng HOG của các đối tượng hiện tại và đặc trưng HOG của các đối tượng trong khung hình quá khứ để dự đoán đặc trưng của đối tượng ở khung hình kế tiếp tương ứng với một hộp bao quanh dự đoán. Khi đến khung hình kế tiếp, các hộp bao quanh dự đoán được so sánh với các hộp bao quanh phát hiện được thông qua ma trận (9).

Từ các cơ sở toán học trên, chúng tôi xây dựng sơ đồ thực hiện bám đối tượng như hình 9. Có thể thấy, sơ đồ bám đối tượng là quá trình lặp lại từ khung hình thứ 2 tới khung hình cuối cùng. Tại bước cuối của chu trình, danh sách đối tượng được bám sẽ được hiển thị hoặc được sử dụng để kiểm tra các điều kiện xử lý khác như phân tích hành vi, đánh giá mức độ nghi vấn,... và sau đó sẽ được lưu ra một tập khung hình riêng biệt để phục vụ các nghiệp vụ chuyên ngành khác.



Hình 9. Sơ đồ giải thuật bám đối tượng

III. THỬ NGHIỆM, ĐÁNH GIÁ

Mỗi giải thuật thực hiện bám đối tượng chúng tôi đã đề xuất đều được thử nghiệm với các dữ liệu tiêu chuẩn để đánh giá được hiệu quả và độ chính xác. Các dữ liệu tiêu chuẩn là tuần tự ảnh của các video được thu có chủ đích trong nhiều điều kiện khác nhau về chiếu sáng, mật độ đối tượng, góc nhìn thay đổi, tỉ lệ đối tượng thay đổi, có sự che khuất giữa các đối tượng, thay đổi độ phân giải,...

Các thử nghiệm về cải thiện tốc độ xử lý được đánh giá trên một máy tính chạy hệ điều hành Windows 10, chip Intel Core i7, tốc độ CPU 2.2 GHz, 8GB RAM, với card màn hình NVIDIA GeForce GTX 1050 Ti 4GB Memory. Các ảnh được thử nghiệm lần lượt chứa các tập số lượng

người, phương tiện giao thông, khuôn mặt và biển số khác nhau như Bảng 3. Bảng 3 cho thấy hiệu quả tốc độ xử lý các tập số lượng người, phương tiện giao thông, khuôn mặt và biển số của mô hình đề xuất. Kiểu ảnh đầu vào đề thử nghiệm gồm có 03 kiểu, kiểu ảnh thứ nhất gồm 48 đối tượng người, 1 đối tượng xe ô tô, 7 đối tượng xe đạp; kiểu ảnh thứ hai gồm 7 đối tượng người; kiểu ảnh thứ 3 gồm 56 khuôn mặt người. Kết quả cho thấy hiệu quả tốc độ tính theo tỷ lệ % của mô hình YOLOv3 đề xuất cao hơn so với mô hình YOLOv3 gốc lần lượt theo từng kiểu ảnh đầu vào là 25%, 20% và 31%.

Bảng 3. Hiệu quả tốc độ xử lý các tập số lượng người, phương tiện giao thông, khuôn mặt và biển số của mô hình đề xuất

Kiểu ảnh đối tượng	YOLO v3 gốc (FPS)	Mô hình AI cải tiến (FPS)	Hiệu quả tốc độ (%)
48 đối tượng người, 1 đối tượng xe ô tô, 7 đối tượng xe đạp 	8.1 fps	10.2 fps	Tăng 25%
7 đối tượng người 	18.1 fps	21.5 fps	Tăng 20%
56 khuôn mặt người 	30.5 fps	40.2 fp	Tăng 31%

Để đánh giá hiệu quả của các mô hình bám đối tượng trong các khung hình, thông thường các nhóm nghiên cứu sử dụng tập dữ liệu MOTChallenge [22], tập dữ liệu KITTI [23], tập dữ liệu HiEve, tập dữ liệu TAO để đánh giá theo từng tiêu chí. Trong bài báo này, chúng tôi sử dụng tập dữ liệu MOTChallenge để đánh giá hiệu quả và độ chính xác của mô hình đề xuất với một số mô hình tốt nhất được công bố tại <https://motchallenge.net> bởi dữ liệu dùng để huấn luyện cho các mô hình bám đối tượng có các đặc tính như đối tượng bị che khuất bởi một hoặc nhiều đối tượng khác, hình ảnh đối tượng bị biến dạng, mặt nền của đối tượng nhiều xáo trộn, chiếu sáng vào đối tượng biến đổi, chuyển động nhanh, khung hình mờ do chuyển động, quay ngoài mặt phẳng, quay trong mặt phẳng, độ phân giải thấp, ngoài tầm nhìn.

Trong tập dữ liệu MOTChallenge, chúng tôi sử dụng hai phương pháp đánh giá là Clear Mot và phương pháp Identity. Để không phụ thuộc vào kiểu tập dữ liệu, chúng tôi sử dụng các tham số TP, TN, FP, FN, MOTA, MOTP, MT, ML, Flag cho phương pháp Clear Mot; các tham số

IDP, IDR, IDF1 cho phương pháp phương pháp Identity [23].

- TP (True Positives): Số đối tượng được xác định là có, và thực tế là có. Giá trị TP càng cao càng tốt.

- TN (True Negatives): Số đối tượng được xác định là không có hoặc mất dấu, và thực tế là không có hoặc mất dấu. TN càng thấp càng tốt.

- FP (False Positives): Số đối tượng được xác định là có, nhưng thực tế không có. FP càng nhỏ càng tốt.

- FN (False Negatives): Số đối tượng được xác định là không có hoặc mất dấu, nhưng thực tế là có. FN càng nhỏ càng tốt.

- MOTA (Multiple Object Tracking Accuracy): Độ chính xác bám nhiều đối tượng, được tính theo các tham số như số lỗi dương FP (False Positives), số lỗi âm FN (False Negatives), số lần thay đổi định danh đối tượng IDSW, và tổng số đối tượng thực GT (Ground Truth). Tiêu chí này cho thấy độ chính xác giữa các đối tượng bám được và các đối tượng đã được dán nhãn. MOTA càng cao càng tốt. Công thức tính MOTA như sau:

$$MOTA = 1 - \frac{FN + FP + IDSW}{GT} \quad (10)$$

- MOTP (Multiple Object Tracking Precision): Độ chính xác bám nhiều đối tượng, được tính theo số đếm các đối tượng phù hợp trong các khung hình $d_{t,i}$ và số đếm các đối tượng được dán nhãn c_t . Tiêu chí này cho thấy độ chính xác của phần nhận dạng đối tượng của giải thuật bám. MOTP càng cao càng tốt. Công thức tính MOTP như sau:

$$MOTP = \frac{\sum_{t,i} d_{t,i}}{\sum_t c_t} \quad (11)$$

- IDP (ID Precision): Độ chính xác định danh đối tượng, được tính theo các tham số như số lỗi dương định danh IDFP (ID False Positives), số định danh đúng IDTP (ID True Positives). IDP càng cao càng tốt. Công thức tính IDP như sau:

$$IDP = \frac{IDTP}{IDTP + IDFP} \quad (12)$$

- IDR (ID Recall): Độ chính xác phát hiện đối tượng, được tính theo các tham số như số lỗi định danh âm IDFN (ID False Negatives), số định danh đúng IDTP. IDR càng cao càng tốt. Công thức tính IDR như sau:

$$IDR = \frac{IDTP}{IDTP + IDFN} \quad (13)$$

- IDF1 (ID F1 Score): Điểm định dạng F1, cho thấy độ chính xác phát hiện đối tượng so với tổng số đối tượng dán nhãn và đối tượng được tính toán. IDF1 càng cao càng tốt. Công thức tính IDF1 như sau:

$$IDF1 = \frac{2}{\frac{1}{IDP} + \frac{1}{IDR}} = \frac{2 \cdot IDTP}{2 \cdot IDTP + IDFP + IDFN} \quad (14)$$

- MT (Mostly Tracked): Tỷ lệ số quỹ đạo bám đúng so với số quỹ đạo được dán nhãn tại ít nhất 80% số khung hình thử nghiệm. MT càng cao càng tốt.

- ML (Mostly Lost): Tỷ lệ số quỹ đạo báo mất so với số quỹ đạo báo mất được dán nhãn tại ít nhất 10% số khung hình thử nghiệm. ML càng nhỏ càng tốt.

- Flag: Tổng số lần các quy đạo bị phân đoạn do việc bám đối tượng tạo ra so với ít nhất 80% số quỹ đạo được dán nhãn. Flag càng nhỏ càng tốt.

Chúng tôi sử dụng phương pháp Clear Mot và Identity để đánh giá mô hình bám đối tượng do chúng tôi đề xuất. Mỗi phương pháp sử dụng 07 tập dữ liệu được đánh số thứ tự Mot16-02, Mot16-04, Mot16-05, Mot16-09, Mot16-10, Mot16-11, Mot16-13 sau đó lấy giá trị trung bình theo từng tiêu chí đánh giá. Các giá trị trung bình này được so sánh với các giá trị tương ứng của 04 mô hình đã được công bố trên mục Public của MOTChallenge tại địa chỉ <https://motchallenge.net/results/MOT16/> ngày 31/5/2024, được thể hiện tại Bảng 4.

Bảng 4. So sánh kết quả thử nghiệm của mô hình đề xuất với các mô hình đã công bố

Phương pháp	Giải thuật	Tiêu chí đánh giá							
		MO TA (%)	MO TP (%)	IDP (%)	IDR (%)	IDF1 (%)	MT (%)	ML (%)	Flag (%)
CLEAR MOT	Mô hình đề xuất	75.6	81.7				304	31	108
	OUTTrack_fm_p	69.2	75.5				283	145	2.6
	MAT	67.6	73.6				188	202	62.3
	Hugmot	64.5	70.5				216	208	3.5
	XJTU	64.4	70.5				220	204	3.6
Identity	Mô hình đề xuất			81.5	72.7	76.8			
	OUTTrack_fm_p			75.5	61.3	67.5			
	MAT			81.2	60.8	69.5			
	Hugmot			75.7	53.5	62.7			
	XJTU			75.9	53.8	62.9			

Trong phương pháp Clear Mot tại Bảng 4 cho thấy, mô hình đề xuất có độ chính xác bám nhiều đối tượng đạt được theo hai tiêu chí về độ chính xác bám đối tượng MOTA và MOTP lần lượt là 75.6% và 81.7% và các giá trị MT, Flag ML lần lượt là 304, 31 và 1.8.

Trong phương pháp Identity tại Bảng 4 chúng tôi đưa ra 3 tiêu chí để đánh giá, so sánh với các mô hình đã được công bố trên mục Public của MOTChallenge cho thấy, mô hình đề xuất có độ chính xác định danh đối tượng IDP, độ chính xác phát hiện đối tượng IDR, IDF1 lần lượt là 81.5, 72.7, 76.8. Các kết quả này cho thấy độ chính xác bám đối tượng, định danh đối tượng và phát hiện đối tượng của mô hình đề xuất tốt hơn so với các mô hình OUTTrack_fm_p, MAT, Hungmot và XJTU đã công bố trước đó.

Thông qua việc so sánh các tiêu chí của hai phương pháp nêu trên cho thấy, mô hình đề xuất trong bài báo này đạt được độ chính xác trong bám, phát hiện và định danh đối tượng từ các khung hình cho kết quả tốt, phù hợp áp dụng cho các hệ thống truy vết đối tượng. Kết quả nghiên cứu này có thể đưa vào ứng dụng trong nhiều lĩnh vực của đời sống xã hội, đặc biệt là trong lĩnh vực an ninh quốc phòng.

IV. KẾT LUẬN

Phương pháp bám đối tượng đã đề xuất đã góp phần giải quyết bài toán hỗ trợ con người trong việc theo dõi các đối tượng bằng “nhìn máy”, giúp giảm công sức và thời gian theo dõi thủ công, tăng cường độ chính xác do không phụ thuộc vào đặc tính sinh học của mắt người theo dõi như nhìn lâu mỏi mắt, mất tập trung, giảm sự chính xác,... hỗ trợ tốt cho công tác điều tra, truy vết đối tượng cần giám sát. Các kết quả đạt được như độ chính xác bám đối tượng MOTA và MOTP lần lượt là 75.6% và 81.7% và các giá trị MT, Flag ML lần lượt là 304, 31 và 1.8; độ chính xác định danh đối tượng IDP, phát hiện đối tượng IDR và IDF1 lần lượt là 81.5%, 72.7% và 76.8%. Có thể thấy mô hình đề xuất cho kết quả khá tốt so với các giải thuật khác ở hạng mục Public của MOTChallenge, nhưng do số lượng dữ liệu huấn luyện còn ở mức hạn chế nên chất lượng chưa được cao. Hướng phát triển tiếp theo sẽ tiếp tục thu thập dữ liệu ở nhiều điều kiện môi trường khác nhau để huấn luyện mô hình trí tuệ nhân tạo cho phát hiện đối tượng hiệu quả hơn, từ đó sẽ nâng cao chất lượng giải thuật bám đối tượng để phù hợp áp dụng cho các hệ thống truy vết đối tượng ứng dụng trong nhiều lĩnh vực đời sống xã hội, đặc biệt là trong lĩnh vực an ninh, quốc phòng.

LỜI CẢM ƠN

Nghiên cứu này được tài trợ bởi Đề tài nghiên cứu khoa học mã số CNC.2021.A06.01.

TÀI LIỆU THAM KHẢO

- [1] Ethem Alpaydin 2014 Introduction to Machine Learning (MIT Press).
- [2] Wold, S.: Principal component analysis. Chemom. Intell. Lab. Syst. 2(1), 37–52 (1987)
- [3] Kim, W., et al.: SVD face: Illumination-invariant face representation. IEEE Signal Process. Lett. 21(11), 1336–1340 (2014).
- [4] Firmanda Al Islama Achyunda Putra, Fitri Utaminigrum, Wayan Firdaus Mahmudy, HOG Feature Extraction and KNN Classification for Detecting Vehicle in The Highway, IJCCS, Vol 14, No 3 (2020).
- [5] Eisha Akanksha, Pamarthi Rama Koteswara Rao, A Feature Extraction Approach for Multi-Object Detection Using HoG and LTP, International Journal of Intelligent Engineering and System, Vol.14, No.5, 2021.
- [6] Melati Tamba, Muhathir, Badriana, Mutammimul Ula, Ilham Sahputra, Classification of Autism Histogram of Oriented Gradient (HOG) Feature Extraction with Support Vector Machine (SVM) Method, 2023 International Conference on Modeling & E-Information Research, Artificial Learning and Digital Applications (ICMERALDA).
- [7] H. Ren and Z. N. Li, "Object detection using edge histogram of oriented gradient", 2014 IEEE International Conference on Image Processing (ICIP), pp. 4057-4061, Oct 2014.
- [8] Boris Babenko, Student Member, Robust Object Tracking with Online Multiple Instance Learning, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 33, NO. 8, AUGUST 2011.
- [9] Helmut Grabner, Michael Grabner, and Horst Bischof. Real-time tracking via on-line boosting. In BMVC, volume 1, page 6, 2006.
- [10] Dattathreya, Sangpil Han, Min-jae Kim, Vivek Maik, Joonki Paik, Keypoint-based object tracking using modified median flow, 2016 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)

- [11] Zdenek Kalal, Krystian Mikolajczyk, and Jiri Matas. Tracking-learning-detection. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 34(7):1409-1422, 2012.
- [12] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista. Exploiting the circulant structure of tracking-by-detection with kernels. In proceedings of the European Conference on Computer Vision, 2012.
- [13] David Held, Sebastian Thrun, and Silvio Savarese. Learning to track at 100 fps with deep regression networks. In European Conference Computer Vision (ECCV), 2016.
- [14] Emre Tunçay, Anil Çelebi, Implementation of MOSSE Object Tracking Algorithm on FPGA with High Level Synthesis Approach, 2020 12th International Conference on Electrical and Electronics Engineering (ELECO).
- [15] Hua Gong; Hui Li; Ke Xu; Yong Zhang, Object Detection Based on Improved YOLOv3-tiny, 2019 Chinese Automation Congress (CAC).
- [16] Navneet Dalal and Bill Triggs, Histograms of Oriented Gradients for Human Detection, 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05).
- [17] Aman Kataria, M. D. Singh, A Review of Data Classification Using K-Nearest Neighbour Algorithm, International Journal of Emerging Technology and Advanced Engineering, Volume 3, Issue 6, June 2013.
- [18] Jeffri M. Llerena, Luis Felipe Zeni, Lucas N. Kristen, Claudio Jung, Gaussian Bounding Boxes and Probabilistic Intersection-over-Union for Object Detection, IEEE Transactions on Image Processing, Vol.33, pp. 671 - 681, 2024.
- [19] H. W. Kuhn, The Hungarian method for the assignment problem, Naval Research Logistics, Vol.52, Issue1, Pages 7-21, February 2005.
- [20] Hilman Fauzi, Uzma Batool, A Three-bar Truss Design using Single-solution Simulated Kalman Filter Optimizer, Journal of Intelligent Manufacturing and Mechatronics, Vol.1, No.2, 2019
- [21] M. Abdullah Azzam, Uzma Batool, Hilman Fauzi, Design of an Helical Spring using Single-solution Simulated Kalman Filter Optimizer, Journal of Intelligent Manufacturing and Mechatronics, Vol.1, No.2, 2019.
- [22] Anton Milan, Laura Leal-Taixe, Ian Reid, Stefan Roth, Konrad Schindler, MOT16: A Benchmark for Multi-Object Tracking, arXiv preprint arXiv:1603.00831, 2016
- [23] Gioele Ciaparrone, Francisco Luque Sánchez, Siham Tabik, Luigi Troiano, Roberto Tagliaferri, Francisco Herrera, Deep learning in video multi-object tracking: a survey, arXiv:1907.12740, November 20, 2019.

PROPOSING SOLUTION FOR OBJECT TRACKING BASED ON IDENTIFICATION USING ARTIFICIAL INTELLIGENCE MODEL AND OBJECT FEATURE EXTRACTION

Abstract: In this article, we will present a proposal for a technique that tracks people and vehicles based on object recognition using an artificial intelligence model, combined with the targets' special extraction method named HOG (Histogram of the orientation gradients) to create some HOG feature vectors, this HOG feature vectors are used as an input for the proposed object tracking algorithm. Then, the object tracking algorithm used the Kalman filters, combined with solving the Hungarian optimal path matrix to maintain tracking objects of the previous frame and the current frame, and predict the objects sticking in the next frame. The proposed algorithm was tested on the MOTChallenge MOT16 data set, compared to two methods including CLEAR MOT using

the criteria, namely MOTA, MOTP, MT, ML, Flag; and Identity uses the IDP, IDR, IDF1 criteria. The experimental results achieved are better than the previously published OUTTrack_fm_p, MAT, Hugmot, XJTU algorithms. The achieved results have object-tracking accuracy according to MOTA, MOTP, MT, and Flag ML criteria of 75.6%, 81.7%, 304, 31 and 1.8 respectively; and according to the IDP, IDR and IDF1 criteria achieved 81.5%, 72.7% and 76.8% respectively.

Keywords: Object tracking, recognition, feature extraction, HOG, MOTChallenge, CLEAR MOT, Identity.



Nguyễn Văn Tài, nhận bằng tốt nghiệp Đại học Giao thông vận tải Hà Nội năm 2006, nhận bằng thạc sĩ Đại học Bách Khoa Hà Nội năm 2010. Anh bắt đầu làm nghiên cứu sinh tại Học viện Công nghệ Bưu chính viễn thông vào năm 2017 và nhận bằng Tiến sĩ Kỹ thuật điện tử vào năm 2022. Hướng nghiên cứu của anh bao gồm ống dẫn sóng nano plasmonic, trường điện từ và truyền sóng và nghiên cứu trong lĩnh vực phòng chống tội phạm sử dụng công nghệ cao.



Lê Xuân Hưng, nhận bằng tốt nghiệp đại học năm 2004 tại Trường Đại học Giao thông vận tải Hà Nội, nhận bằng thạc sĩ năm 2020 tại Đại học Bách Khoa Hà Nội. Hướng nghiên cứu của anh bao gồm thông tin quang, xử lý dữ liệu, học máy, trí tuệ nhân tạo.