



Bài viết

Bạo lực học đường DT–SVM dựa trên video Thuật toán phát hiện

Liang Ye 1,2,* , Le Wang 1,3, Hany Ferdinando 2,4 , Tapio Seppänen 5 và Esko Alasaarela 2

¹ Khoa Kỹ thuật Thông tin và Truyền thông, Học viện Công nghệ Cáp Nhĩ Tân, Cáp Nhĩ Tân 150080, Trung Quốc; wangle68@huawei.com

² Đơn vị Nghiên cứu OPEM, Đại học Oulu, 90014 Oulu, Phần Lan; Hany.Ferdinando@oulu.fi (HF); Esko.Alasaarela@oulu.fi (EA)

³ Viện Huawei Bắc Kinh, 100085 Bắc Kinh, Trung Quốc

⁴ Khoa Kỹ thuật Điện, Đại học Petra Christian, Surabaya 60236, Indonesia

Nhóm Phân tích Tín hiệu Sinh lý, Đại học Oulu, 90014 Oulu, Phần Lan; tapio.seppanen@oulu.fi *Thư từ: yeliang@hit.edu.cn



Nhận: 3 tháng 3 năm 2020; Được chấp nhận: 1 Tháng Tư 2020; Được phát hành: 3 Tháng Tư 2020

Tóm tắt: **Bắt nạt học đường** là một vấn đề nghiêm trọng ở thanh thiếu niên. **Bạo lực học đường** là một loại bắt nạt học đường và được coi là có hại nhất. Khi các kỹ thuật AI (Trí tuệ nhân tạo) phát triển, hiện có các phương pháp mới để phát hiện bạo lực học đường. Bài báo này đề xuất một thuật toán phát hiện bạo lực học đường dựa trên video. Thuật toán này đầu tiên phát hiện các mục tiêu chuyển động ở tiền cảnh thông qua phương pháp KNN (K-Nearest Neighbor) và sau đó xử lý trước các mục tiêu được phát hiện thông qua các phương pháp xử lý hình thái. Sau đó, bài báo này đề xuất một phương pháp tích hợp khung hình chữ nhật có giới hạn để tối ưu hóa khung hình chữ nhật có giới hạn của các mục tiêu chuyển động. Các đặc điểm khung hình chữ nhật và các đặc điểm dòng chảy quang học đã được trích xuất để mô tả sự khác biệt giữa bạo lực học đường và các hoạt động trong cuộc sống hàng ngày.

Chúng tôi đã sử dụng thuật toán Relief-F và Wrapper để giảm kích thước đối tượng. SVM (Máy vector hỗ trợ) đã được áp dụng làm bộ phân loại và xác thực chéo gấp 5 lần đã được thực hiện. Độ chính xác là 89,6% và độ chính xác là 94,4%. Để cải thiện hơn nữa hiệu suất nhận dạng, chúng tôi đã phát triển bộ phân loại hai lớp DT–SVM (Decision Tree–SVM). Chúng tôi đã sử dụng biểu đồ hộp để xác định một số đặc điểm của lớp DT có thể phân biệt giữa bạo lực thể chất diễn hình và các hoạt động trong cuộc sống hàng ngày và giữa các hoạt động diễn hình trong cuộc sống hàng ngày và bạo lực thể chất. Đối với các hoạt động còn lại, lớp SVM thực hiện phân loại. Đối với bộ phân loại DT–SVM này, độ chính xác đạt 97,6% và độ chính xác đạt 97,2%, do đó cho thấy sự cải thiện đáng kể.

Từ khóa: nhận dạng hoạt động; xử lý hình ảnh; nhận dạng mẫu; Phát hiện bạo lực học đường

1. Giới thiệu

Trong thế giới hiện đại, internet đã được tích hợp vào cuộc sống hàng ngày của con người. Trong khi thông tin trực tuyến làm phong phú thêm cuộc sống của mọi người, xác suất trẻ em tiếp xúc với nội dung bạo lực và đẫm máu cũng tăng lên trong môi trường này. Kết quả là, bắt nạt học đường ngày càng trở nên phổ biến. Bắt nạt học đường là một sự kiện bạo lực xảy ra dưới nhiều hình thức khác nhau, chẳng hạn như bạo lực thể chất, bắt nạt bằng lời nói, phá hoại tài sản cá nhân, v.v. Bạo lực thể chất được coi là có hại nhất đối với thanh thiếu niên. Theo "Bạo lực và bắt nạt trong khuôn viên trường" do UNESCO (Tổ chức Giáo dục, Khoa học và Văn hóa Liên hợp quốc) công bố vào năm 2017, 32,5% sinh viên trên toàn thế giới bị bắt nạt trong khuôn viên trường, với tổng số 243 triệu. Vì vậy, phòng chống bắt nạt học đường là một chủ đề cấp bách nhưng vượt thời gian.

Các nghiên cứu về phòng chống bắt nạt học đường đã được phát triển từ những năm 1960. Các phương pháp ngăn chặn bắt nạt học đường từng do con người điều khiển, tức là khi một sự kiện bắt nạt học đường xảy ra,

Những người ngoài cuộc sẽ báo cáo sự kiện cho các giáo viên. Tuy nhiên, những người ngoài cuộc có thể sợ bị trả thù từ những kẻ bắt nạt của họ và do đó thường không báo cáo sự kiện. Khi điện thoại thông minh ngày càng trở nên phổ biến, các ứng dụng chống bắt nạt đã được phát triển để nạn nhân sử dụng. Tuy nhiên, các ứng dụng này cũng do con người điều khiển. Khi bắt nạt học đường xảy ra, nạn nhân vận hành ứng dụng để gửi tin nhắn analarm, điều này có thể khiến kẻ bắt nạt tức giận gây hại thêm.

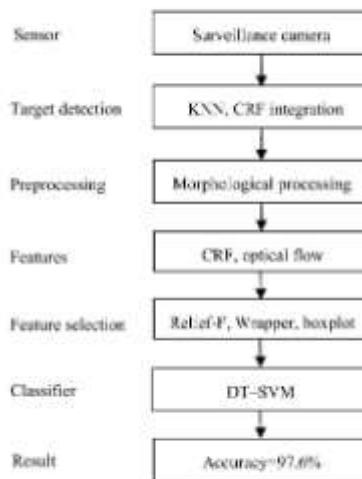
Khi các kỹ thuật trí tuệ nhân tạo phát triển, các phương pháp tự động đang được phát triển để phát hiện bắt nạt học đường. Trong nghiên cứu trước đây của chúng tôi [1–4], chúng tôi đã sử dụng cảm biến chuyển động để phát hiện bạo lực thể chất. Các cảm biến chuyển động này thu thập dữ liệu gia tốc 3D và dữ liệu con quay hồi chuyển 3D từ người dùng. Sau đó, các thuật toán trích xuất các tính năng miền thời gian và các tính năng miền tần số từ dữ liệu đó. Các thuật toán lựa chọn tính năng, chẳng hạn như Relief-F [5] và Wrapper [6], được sử dụng để loại trừ các tính năng vô dụng. PCA (Phân tích thành phần chính) [7] và LDA (Phân tích phân biệt tuyến tính) [8] tiếp tục làm giảm chiều tính năng. Chúng tôi đã phát triển một số bộ phân loại cho các tính năng khác nhau và các hoạt động khác nhau, chẳng hạn như FMT (Muzzy Multi-Thresholds) [1], PKNN (Proportional K-Nearest Neighbor) [2], BPNN (BackPropagation Neural Network) [3] và DT-RBF (Decision Tree–Radial Basis Function) [4], cuối cùng đạt được độ chính xác trung bình là 93,7%. Các thuật toán này phát hiện bạo lực thể chất từ quan điểm của nạn nhân. Trong trường hợp những kẻ bắt nạt loại bỏ các cảm biến chuyển động khỏi nạn nhân, bài báo này đề xuất một phương pháp phát hiện bắt nạt học đường thay thế dựa trên camera giám sát trong khuôn viên trường. Hình 1 minh họa cấu trúc của hệ thống phát hiện bạo lực này.



Hình 1. Cấu trúc của hệ thống phát hiện bạo lực.

Camera giám sát được sử dụng là camera an ninh thông thường chỉ chụp hình ảnh của khuôn viên trường. Tất cả các thủ tục nhận dạng được thực hiện bằng máy tính. Khi một sự kiện bạo lực đã được phát hiện, cảnh báo sẽ được gửi đến giáo viên và / hoặc phụ huynh bằng tin nhắn ngắn hoặc các phương tiện truyền thông xã hội khác. Hình 2 cho thấy sơ đồ của phương pháp phát hiện được đề xuất.

Có thể có một số camera trong khuôn viên trường, mỗi camera giám sát một khu vực nhất định. Camera chụp ảnh khu vực giám sát và đầu tiên phát hiện xem có mục tiêu di chuyển hay không. Nếu có mục tiêu di chuyển, thuật toán KNN (K-Nearest Neighbors) sau đó trích xuất các mục tiêu tiềm cảnh. Chúng tôi sử dụng các phương pháp xử lý hình thái để xử lý trước các mục tiêu được phát hiện và đề xuất phương pháp tích hợp khung hình chữ nhật để tối ưu hóa mục tiêu di chuyển được phát hiện. Theo sự khác biệt giữa bạo lực thể chất và các hoạt động trong cuộc sống hàng ngày, chúng tôi trích xuất các tính năng khung hình chữ nhật giới hạn, chẳng hạn như tỷ lệ khung hình và các tính năng dòng chảy quang học, sau đó giảm kích thước tính năng với Relief-F và Wrapper. Bằng cách vẽ hộp, chúng tôi xác định rằng một số đặc điểm có thể phân biệt chính xác bạo lực thể chất với các hoạt động hàng ngày, vì vậy chúng tôi đã thiết kế bộ phân loại hai lớp DT-SVM. Lớp đầu tiên là Cây quyết định tận dụng các tính năng như vội và lớp thứ hai là SVM sử dụng các tính năng còn lại để phân loại. Theo kết quả mô phỏng, độ chính xác đạt 97,6% và độ chính xác đạt 97,2%.



Hình 2. Sơ đồ của phương pháp phát hiện bạo lực được đề xuất (KNN, K-Nearest Neighbors; CRF, khung hình chữ nhật có giới hạn; và DT – SVM, Cây quyết định – Máy vectơ hỗ trợ).

Phần còn lại của bài báo này được tổ chức như sau: Phần 2 khám phá một số công việc liên quan nhận dạng hoạt động dựa trên video, Phần 3 cho thấy các quy trình thu thập dữ liệu và tiền xử lý dữ liệu, Phần 4 mô tả các phương pháp trích xuất tính năng và lựa chọn tính năng, Phần 5 xây dựng bộ phân loại DT-SVM 2 lớp, Phần 6 phân tích kết quả mô phỏng và Phần 7 trình bày kết luận.

2. Nhận dạng hoạt động dựa trên video

Nhận dạng hoạt động là một chủ đề phổ biến trong các lĩnh vực trí tuệ nhân tạo [9,10] và thành phố thông minh [11,12]. Hầu hết các nghiên cứu liên quan tập trung vào hoạt động sinh hoạt hàng ngày hoặc công nhận thể thao. Sun et al. [13] đã nghiên cứu nhận dạng hành động dựa trên biểu diễn động học của dữ liệu video. Các tác giả đã đề xuất một bộ mô tả động học có tên là Vận tốc tính năng tĩnh và động (SDEV), mô hình hóa sự thay đổi của cả thông tin tĩnh và động với thời gian nhận dạng hành động. Họ đã thử nghiệm thuật toán của mình trên một số hoạt động thể thao và đạt được độ chính xác trung bình lần lượt là 89,47% đối với các môn thể thao UCF (Đại học Trung tâm Florida) và 87,82% đối với các môn thể thao Olympic. Wang et al. [14] đã nghiên cứu nhận dạng hành động bằng cách sử dụng biểu diễn thành phần hành động không tiêu cực và lựa chọn cơ sở thưa thớt. Các tác giả đã đề xuất mô tả thời gian không gian nhận biết ngữ cảnh, các đơn vị học hành động sử dụng phân tích ma trận không âm được chính quy hóa đồ thị và một mô hình thưa thớt. Đối với bộ dữ liệu thể thao của UCF, họ đạt được độ chính xác trung bình là 88,7%. Tu et al. [15] đã nghiên cứu VLAD không gian thời gian nhấn mạnh giai đoạn hành động để nhận dạng hành động video. Các tác giả đã đề xuất một phương pháp vectơ không gian thời gian nhấn mạnh giai đoạn hành động của các mô tả tổng hợp cục bộ (Actions-ST-VLAD) để tổng hợp các tính năng sâu thông tin trên toàn bộ video theo phân đoạn tính năng video thích ứng và lấy mẫu tính năng phân đoạn thích ứng (AVFS-ASFS). Hơn nữa, họ đã khai thác phương thức RGB để chụp các vùng nổi bật chuyển động trong hình ảnh RGB tương ứng với các hoạt động hành động. Đối với bộ dữ liệu UCF101, họ đạt được độ chính xác trung bình là 97,9%, đối với bộ dữ liệu HMDB51 (cơ sở dữ liệu chuyển động của con người), họ đạt được 80,9% và đối với hoạt động rộng họ đạt được 90,0%.

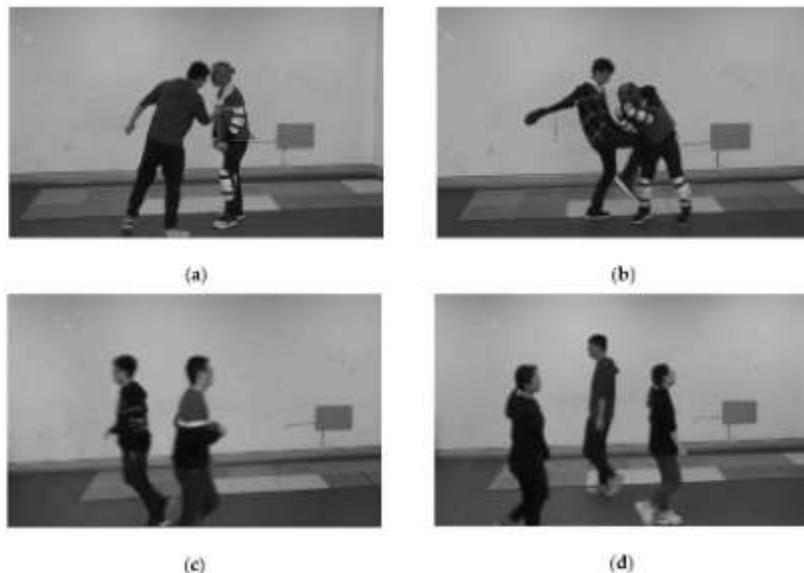
Trong những năm gần đây, các nhà nghiên cứu đã bắt đầu chú ý đến việc phát hiện bạo lực. AS Keçeli và cộng sự [16] đã nghiên cứu phát hiện hoạt động bạo lực bằng phương pháp học chuyển giao. Các tác giả đã thiết kế một máy dò bạo lực dựa trên chuyển giao và thử nghiệm nó trên ba bộ dữ liệu. Đối với dòng chảy bạo lực (video được tải xuống từ YouTube), họ đạt được độ chính xác trung bình là 80,90%, đối với khúc côn cầu, họ đạt được 94,40% và đối với phim, họ đạt được 96,50%. Ha et al. [17] nghiên cứu phát hiện bạo lực cho hệ thống giám sát video

sử dụng thông tin chuyển động bát thường. Các tác giả đã ước tính vectơ chuyển động bằng cách sử dụng phương pháp tiếp cận CombinedLocal-Global với Total Variation (CLG-TV) sau khi phát hiện mục tiêu và phát hiện các sự kiện bạo lực bằng cách phân tích các đặc điểm của các vectơ chuyển động được tạo ra trong vùng của đối tượng bằng cách sử dụng Tính năng đồng xuất hiện chuyển động (MCF). Họ đã sử dụng cơ sở dữ liệu CAVIAR nhưng không đưa ra kết quả ước tính cho độ chính xác trung bình. Tahereh Zarrat Ehsan và cộng sự [18] đã nghiên cứu phát hiện bạo lực trong camera giám sát trong nhà bằng cách sử dụng quỹ đạo chuyển động và biểu đồ vi sai của luồng quang học. Họ trích xuất quỹ đạo chuyển động và các đặc điểm không gian và sau đó sử dụng SVM để phân loại. Họ cũng sử dụng bộ dữ liệu CAVIAR và đạt được độ chính xác trung bình là 91%.

Bài báo này nghiên cứu việc phát hiện bạo lực học đường, một loại bạo lực thể chất xảy ra trong khuôn viên trường. Các phần sau đây sẽ giải thích chi tiết tập dữ liệu được sử dụng và các thuật toán được đề xuất.

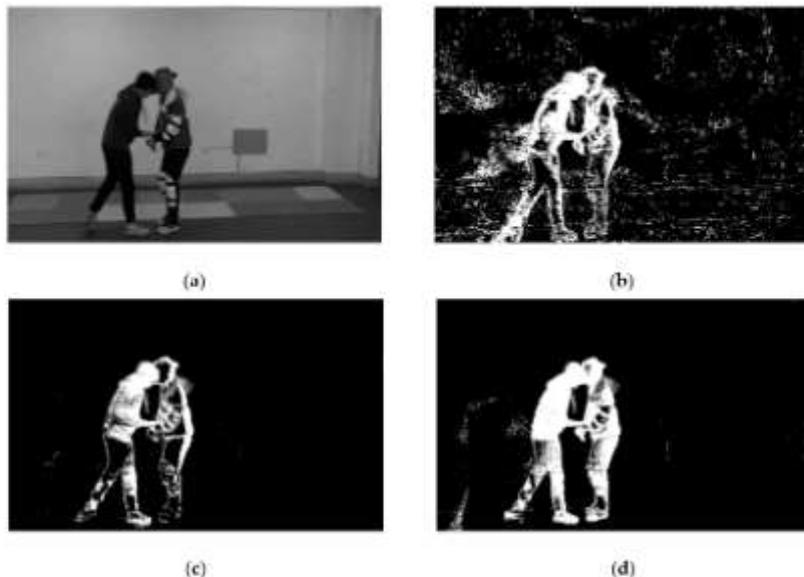
3. Thu thập dữ liệu và tiền xử lý dữ liệu

3.1. Thu thập dữ liệu Bạo lực học đường khác với bạo lực xã hội, chẳng hạn như đánh nhau trên đường phố hoặc đấm bốc, ở những điểm sau: (1) nạn nhân thường không chống cự, (2) không sử dụng vũ khí và (3) thanh thiếu niên không mạnh bằng người lớn, vì vậy biên độ chuyển động cơ thể của nạn nhân không lớn bằng trong một cuộc chiến xã hội giữa người lớn. Vì hiện tại không có bộ dữ liệu bạo lực học đường công lập, chúng tôi đã thu thập dữ liệu bằng cách nhập vai bạo lực học đường và các hoạt động hàng ngày, bao gồm cả thể thao trong khuôn viên trường. Dữ liệu được thu thập bằng cách nhập vai bạo lực học đường và các hoạt động sinh hoạt hàng ngày. Một số tình nguyện viên thay phiên nhau đóng vai những kẻ bắt nạt và bị bắt nạt. Kẻ bắt nạt mặc đồ bảo hộ để tránh bị thương bất ngờ. Tổng cộng, 24.896 khung hình đã được ghi lại, bao gồm 12.448 khung hình bạo lực học đường, 9963 khung hình hoạt động cuộc sống hàng ngày và 2485 khung hình tĩnh. Hình 3 cho thấy một số ví dụ về bạo lực học đường và các hoạt động trong cuộc sống hàng ngày.



Hình 3. Một số ví dụ về bạo lực học đường và các hoạt động sinh hoạt hàng ngày: (a) đấm; (b) đá; (c) hai người chạy; và (d) ba người đi bộ.

3.2. Phát hiện mục tiêu và tiền xử lý Có một số phương pháp để phát hiện mục tiêu tiền cảnh, chẳng hạn như sử dụng vi sai hoặc khung hình vi sai tĩnh [19]. Vi sai tĩnh là một phương pháp thường được sử dụng để phát hiện mục tiêu di chuyển. Đầu tiên, nó lưu trữ một hình nền và sau đó so sánh từng ảnh đầu vào với nền này. Mục tiêu chuyển động được phát hiện theo sự khác biệt giữa hình ảnh đầu vào và ảnh nền. Vi sai tĩnh có một nhược điểm đáng kể do các biến thể trong nền, chẳng hạn như sự thay đổi của ánh sáng. Bạo lực học đường thường xảy ra ngoài trời, nơi ánh sáng của khung cảnh thay đổi vào ban ngày, vì vậy Vi sai tĩnh không phù hợp cho mục đích này. Khung vi sai hoạt động theo cách tương tự nhưng so sánh khung hiện tại với khung trước đó. Do đó, nó mạnh hơn khi chống lại sự thay đổi của điều kiện ánh sáng so với chênh lệch tĩnh, nhưng cả hai mục tiêu trên khung hình hiện tại và trên khung hình cuối cùng sẽ xuất hiện trong kết quả được phát hiện. Một cải tiến là nhân kết quả được phát hiện với khung hiện tại, nhưng điều này sẽ làm tăng lỗi phát hiện. Khi các kỹ thuật học máy phát triển, hiện có các phương pháp mới để phát hiện mục tiêu tiền cảnh, chẳng hạn như KNN (K-Nearest Neighbors) [20], MOG2 (Hỗn hợp Gaussians) [21] và GMG (Geometric Multigrid) [22]. KNN thường được sử dụng như một bộ phân loại. KNN phân loại mẫu thử nghiệm theo khoảng cách Euclid của nó đến K mẫu đào tạo gần nhất trong mỗi lớp. Để phát hiện tiền cảnh, KNN so sánh sự giống nhau của các pixel. MOG2 dựa trên Mô hình hỗn hợp Gaussian (GMM). MOG2 chọn một lượng phân phối Gaussian thích hợp cho mỗi pixel, vì vậy nó cung cấp các điều chỉnh tốt hơn cho các thay đổi cảnh. GMG kết hợp ước tính hình nền tĩnh và phân đoạn Bayes của từng pixel. Nó mô hình hóa nền với các khung trước đó và xác định các mục tiêu tiền cảnh bằng ước tính Bayes. Vì vi sai tĩnh và vi sai khung hình có những nhược điểm nêu trên, chúng tôi đã chọn các thuật toán KNN, MOG2 và GMG để trích xuất các mục tiêu tiền cảnh. Kết quả phát hiện mục tiêu tiền cảnh được đưa ra trong Hình 4.



Hình 4. Trích xuất mục tiêu tiền cảnh với các thuật toán khác nhau: (a) hình ảnh gốc; (b) Phát hiện GeometricMultigrid (GMG); (c) Phát hiện hỗn hợp Gaussian (MOG2); và (d) Phát hiện K-Hàng xóm gần nhất (KNN).

Hình 4a là hình ảnh gốc do máy ảnh chụp. Hình 4b là phát hiện mục tiêu của GMG. Có quá nhiều nhiễu trong bức ảnh, và đường viền của người bên trái không rõ ràng. Hình 4cis phát hiện mục tiêu bằng MOG2. Mặc dù nó có ít nhiễu nhất trong số ba thuật toán, nhưng hầu hết các pixel của người ở phía bên phải đều bị nhầm là nền. Hình 4d minh họa phát hiện mục tiêu bằng KNN. KNN đạt được sự cân bằng tốt nhất giữa thông tin tiền cảnh và tiếng ồn xung quanh. Do đó, bài báo này sử dụng KNN để trích xuất các mục tiêu tiền cảnh. Đối với tiếng ồn xung quanh, chúng tôi sử dụng bộ lọc trung bình. Sau đó, nhị phân hóa được thực hiện để tăng cường các mục tiêu tiền cảnh.

3.3. Xử lý hình thái Để có được đường viền của các mục tiêu, chúng tôi đã thực hiện xử lý hình thái. Thông thường, quá trình xử lý hình thái bao gồm giãn nở, xói mòn, hoạt động mở và hoạt động đóng. Giãn nở và xói mòn là hai hoạt động cơ bản trong quá trình xử lý hình thái. Giãn nở thường được sử dụng để kết nối các phần riêng biệt trong hình ảnh, trong khi xói mòn thường được sử dụng để giảm nhiễu. Tuy nhiên, để tránh thay đổi hình dạng của mục tiêu tiền cảnh, chúng thường hoạt động theo cặp, tạo thành hoạt động mở và đóng. Hoạt động mở tương tự như xói mòn sau đó là giãn nở và được sử dụng để loại bỏ các điểm nhiễu xung quanh lớn và các cạnh nhẵn, trong khi hoạt động đóng giống như giãn nở sau đó là xói mòn và được sử dụng để kết nối các bộ phận riêng biệt trong hình ảnh và lắp đầy các lỗ và khoảng trống do giảm nhiễu gây ra. Bằng cách chọn các hoạt động xử lý hình thái thích hợp, người ta có thể có được các mục tiêu tiền cảnh hoàn chỉnh với ít nhiễu nhất. Hình 5 cho thấy chuỗi xử lý hình ảnh. Hình 5a là hình ảnh gốc và Hình 5b là hình ảnh sau khi phát hiện mục tiêu tiền cảnh, lọc trung bình và nhị phân hóa. Trong Hình 5b, có một nhiễu khối ở góc dưới bên trái do hiệu ứng bóng. Để loại bỏ nhiễu khối này, trước tiên chúng tôi áp dụng thao tác mở. Như thể hiện trong Hình 5c, diện tích của nhiễu khối đã giảm, nhưng các mục tiêu tiền cảnh hầu như không bị ảnh hưởng. Sau đó, xói mòn được áp dụng để loại bỏ nhiễu này. Tiếp theo, chúng tôi thực hiện một thao tác đóng, sau đó bàn chân của người ở phía bên phải được kết nối với chân của anh ta và các lỗ trên cánh tay của anh ta được lắp đầy, như thể hiện trong Hình 5e. Cuối cùng, sự giãn nở được thực hiện để kết nối chân của người ở phía bên trái, như thể hiện trong Hình 5f. Mặc dù các mục tiêu được trích xuất là dày hơn người thật, nhưng phác thảo này có thể thể hiện hoạt động của những người đó.

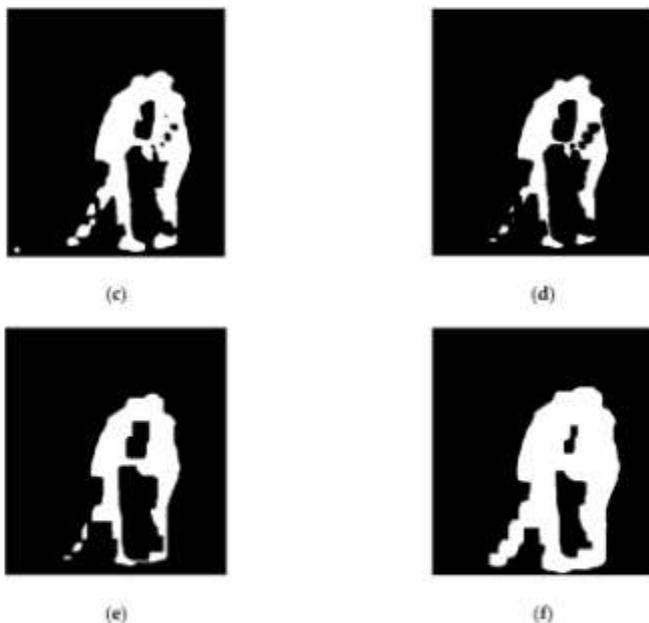


(a)



(b)

Hình 5. Cont.



Hình 5. Xử lý hình thái: (a) hình ảnh gốc; (b) trích xuất mục tiêu và nhị phân hóa; (c) hoạt động mờ; (d) xói mòn; (e) đóng hoạt động; và (f) giãn nở.

3.4. Tích hợp khung hình chữ nhật giới hạn Vì đường viền của các mục tiêu tiền cảnh quá phức tạp đối với máy tính để phân tích và vì thuật toán phát hiện bạo lực học đường thời gian thực phải có độ phức tạp tính toán thấp, khung hình chữ nhật có giới hạn thường được sử dụng thay vì chính phác thảo. Sau khi phát hiện mục tiêu tiền cảnh được đề cập trong Phần 3.3, chúng tôi có được vị trí của các điểm ảnh mục tiêu tiền cảnh và tìm thấy các mục tiêu trong ảnh gốc. Tuy nhiên, phát hiện mục tiêu tiền cảnh không phải lúc nào cũng chính xác mặc dù xử lý hình thái đã được thực hiện. Hình 6 đưa ra hai ví dụ về khung hình chữ nhật có giới hạn bất ngờ.



Hình 6. Hai ví dụ về khung hình chữ nhật có giới hạn bất ngờ: (a) một phần tách ra khỏi toàn bộ và (b) khung hình chữ nhật có giới hạn dư thừa.

Trong Hình 6a, chân phải của người bên trái được tách ra khỏi cơ thể của anh ta trong quá trình phát hiện mục tiêu tiền cảnh, và chân trái bị mất, vì vậy có hai khung hình chữ nhật được giới hạn cho người này. Khung lớn hơn bao phủ người đó từ đầu đến chân, trong khi khung nhỏ hơn che chân phải của người đó. Hình 6b cho thấy một tình huống khác. Một chân của người ở phía bên trái được tách ra khỏi cơ thể của anh ta trong quá trình phát hiện mục tiêu tiền cảnh, nhưng phần chân còn lại được kết nối với nhau, vì vậy cũng có hai khung hình chữ nhật được giới hạn cho người này. Khung lớn hơn bao phủ anh ta cùng với người kia vì họ đã chạm vào nhau, trong khi khung nhỏ hơn che chân phải của anh ta. Ngoài ra còn có một khung hình chữ nhật được giới hạn thứ ba che một ổ cảm trên tường trong quá trình phát hiện mục tiêu tiền cảnh do nơi trú ẩn khi người đó di chuyển. Để có được khung hình chữ nhật có giới hạn chính xác cho các mục tiêu tiền cảnh, các khung dự phòng như vậy nên được cố định.

Chúng tôi đề xuất một phương pháp tích hợp khung hình chữ nhật có giới hạn. Thuật toán này là phát hiện va chạm dựa trên. Trong cảnh 2-D, va chạm được đánh giá bằng các khung hình chữ nhật được giới hạn của các mục tiêu tiền cảnh. Nếu khoảng cách giữa hai cạnh bất kỳ của hai khung hình chữ nhật có giới hạn trở thành 0, thì xảy ra va chạm. Trong thuật toán được đề xuất, nếu hai khung hình chữ nhật có giới hạn va chạm theo hướng ngang và gần nhau theo hướng thẳng đứng (nói cách khác, khung nhỏ hơn nằm trong phạm vi di chuyển đều đặn của khung lớn hơn), thì hai khung có thể được tích hợp thành một khung toàn bộ. Giả sử rằng tọa độ của hai khung hình chữ nhật giới hạn là $(x_1, y_1, x_1 + w_1, y_1 + h_1)$ và $(x_2, y_2, x_2 + w_2, y_2 + h_2)$, tương ứng, trong đó w_i là chiều rộng của khung, h_i là chiều cao, (x_i, y_i) là góc dưới bên trái của khung và $(x_i + w_i, y_i + h_i)$ là góc trên bên phải, $i = 1, 2$. Hai khung hình chữ nhật có giới hạn chỉ có thể được tích hợp khi

$$\| (x_1 + v_{12}) - (x_2 + v_{22}) \| < \| v_1 + v_2 \| \quad (1)$$

và

$$\| (Y_1 + H_{12}) - (Y_2 + H_{22}) \| + h_{th} < \| H_1 + H_2 \| \quad (2)$$

trong đó $h_{th} > 0$ là ngưỡng dung sai thể hiện sự khác biệt vị trí của hai khung theo hướng thẳng đứng.

Khi tích hợp hai khung hình chữ nhật có giới hạn, tọa độ ngoài cùng của hai khung trở thành tọa độ của khung hình chữ nhật có giới hạn tích hợp, tức là $(\min\{x_i\}, \min\{y_i\}, \max\{x_i + w_i\} \text{ và } \max\{y_i + h_i\})$, $i = 1, 2$. Hình 7 cho thấy các hình ảnh cố định theo tích hợp khung hình chữ nhật giới hạn.



(a)



(b)

Hình 7. Hình ảnh sau khi tích hợp khung hình chữ nhật giới hạn: (a) cố định Hình 4a và (b) cố định Hình 4b.

So sánh Hình 7a với Hình 6a, chân phải của người ở phía bên trái được kết nối lại với cơ thể của anh ta và hai mục tiêu tiền cảnh đều được phát hiện chính xác. So sánh Hình 7b với Hình 6b, hai khung hình chữ nhật nhỏ có giới hạn dự phòng trên chân và trên ống cắm được tích hợp vào khung lớn. Chỉ có một khung hình chữ nhật được giới hạn trong hình ảnh này bởi vì hai người đến với nhau. Do đó, các khung hình chữ nhật được giới hạn đã được trích xuất chính xác bằng thuật toán được đề xuất.

4. Trích xuất tính năng và lựa chọn tính năng

4.1. Trích xuất tính năng Bây giờ các mục tiêu tiền cảnh (khung hình chữ nhật được giới hạn) đã được phát hiện, các đặc điểm được trích xuất từ các mục tiêu được phát hiện để mô tả sự khác biệt giữa bạo lực thể chất và các hoạt động trong cuộc sống hàng ngày.

4.1.1. Đặc điểm khung hình chữ nhật có giới hạn Các đặc điểm sau đây được trích xuất từ các khung hình chữ nhật có giới hạn được phát hiện.

(1) Số lượng khung hình chữ nhật được giới hạn và sự thay đổi của nó. Thứ nhất, số lượng khung hình chữ nhật được giới hạn có thể phản ánh có bao nhiêu người trong khu vực giám sát. Nếu chỉ có một khung hình hoặc không có khung hình, thì bức ảnh này rất có thể là một cảnh không bạo lực. Sự thay đổi về số lượng khung hình cũng có thể cho biết loại hoạt động trong cảnh. Ví dụ, ban đầu có hai khung hình trong hình ảnh và trong hình ảnh tiếp theo, hai khung hình nối thành một. Điều này có thể có nghĩa là hai người đã gặp nhau và thuật toán sẽ đánh giá liệu bạo lực thể chất có khả năng xảy ra hay không. Ngược lại, nếu có một khung hình lúc đầu sau đó tách thành hai và không bao giờ nối lại, thì đó có lẽ là cảnh không bạo lực. (2) Chiều rộng của khung hình chữ nhật giới hạn và sự thay đổi của nó. Chiều rộng của một giới hạn

Khung hình chữ nhật có thể được sử dụng để đánh giá xem hình ảnh mô tả một người hay nhiều người. Nếu nó cho thấy một người duy nhất, hình ảnh đó có thể là một cảnh bắt bạo động; nếu không, nhiều người đã chạm vào nhau và thuật toán sẽ đánh giá xem nó có đang quan sát một sự kiện bạo lực thể chất hay không. Sự thay đổi của chiều rộng cũng có thể chỉ ra hành động đang diễn ra. Ví dụ, khi một người đánh hoặc đẩy người khác, chiều rộng của khung hình chữ nhật được giới hạn của họ (được phát hiện là một vì họ đã chạm vào nhau) sẽ thay đổi. (3) Chiều cao của các khung hình chữ nhật có giới hạn và sự thay đổi của nó. Chiều cao của một giới hạn

Khung hình chữ nhật có thể phản ánh tư thế của một người. Ví dụ, chiều cao của khung khi người đứng khác với chiều cao khi anh ta ngồi xổm xuống. Sự thay đổi chiều cao có thể phản ánh hành động của một người. Khi có nhiều hơn một khung hình chữ nhật được giới hạn trong bức tranh, sự thay đổi chiều cao có thể cung cấp thêm hướng dẫn. Ví dụ, nếu có hai khung hình trong bức tranh và một trong số chúng đột nhiên trở nên thấp hơn, điều gì đó có thể đã xảy ra, mà thuật toán phát hiện bạo lực thể chất sẽ nhận thấy. (4) Tỷ lệ khung hình của các khung hình chữ nhật có giới hạn và sự thay đổi của nó. Những tính năng này là một

sự kết hợp giữa chiều rộng và chiều cao. Tỷ lệ khung hình phản ánh tư thế của một người và sự biến đổi của nó phản ánh hành động. Khi có nhiều khung hình trong một bức tranh, những đặc điểm này có thể được sử dụng để đánh giá liệu một sự kiện bạo lực thể chất có xảy ra hay không. (5) Diện tích của khung hình chữ nhật giới hạn và sự thay đổi của nó. Các tính năng này có thể được sử dụng để đánh giá số lượng người trong một khung hình chữ nhật giới hạn và loại trừ các mục tiêu di chuyển không mong muốn. Trong một cảnh bạo lực thể chất, các hành động thường căng thẳng, và diện tích của khung hình chữ nhật được giới hạn thay đổi đáng kể và thường xuyên; Do đó, những tính năng này có thể chỉ ra bạo lực trong thuật toán.

(6) Khoảng cách tâm của các khung hình chữ nhật có giới hạn và sự thay đổi của nó. Trong hình ảnh nhị phân sau khi phát hiện mục tiêu tiền cảnh, tâm có thể được coi là trung tâm của mục tiêu. Sau đó, khoảng cách tâm của các khung hình chữ nhật được giới hạn đại diện cho khoảng cách giữa hai người. Khi khoảng cách trung tâm lớn, hai người tương đối xa nhau; Khi khoảng cách trung tâm nhỏ, hai người ở gần nhau và điều gì đó có thể xảy ra giữa họ. (7) Khu vực của các mục tiêu được phát hiện và sự thay đổi của nó. Những đặc điểm này được trích xuất không phải từ giới hạn

khung hình chữ nhật nhưng từ các mục tiêu được phát hiện sau khi xử lý hình thái, như trong Hình 3f. Khi có nhiều mục tiêu (khung hình chữ nhật được giới hạn) trong một hình ảnh, hãy trích xuất giá trị lớn nhất của các đối tượng, ví dụ: chiều rộng tối đa của khung hình chữ nhật có giới hạn. Tuy nhiên, diện tích tối đa và tổng diện tích đều được trích xuất. Vì những đặc điểm này bị ảnh hưởng bởi khoảng cách giữa máy ảnh và những người được quan sát, công việc này không tính đến những người ở xa.

Rõ ràng, cần có một mối quan hệ logic giữa khung hiện tại và khung trước đó. Do đó, bên cạnh các đặc điểm nêu trên được trích xuất trực tiếp từ các khung hình chữ nhật có giới hạn, chúng tôi cũng xác định một số trạng thái cho các khung hình chữ nhật có giới hạn.

Giả sử rằng Areamax là diện tích tối đa của các mục tiêu được phát hiện trong hình ảnh, Areamin là diện tích tối thiểu của các mục tiêu được phát hiện trong hình ảnh, Widthmax là chiều rộng tối đa của các khung hình chữ nhật có giới hạn, AreaThlow là ngưỡng thấp hơn của một người, NonHuth là ngưỡng của không phải con người, AreaThup là ngưỡng trên của một người, MulPer1th là ngưỡng của nhiều người, MulPer2th là một ngưỡng khác của nhiều người, Cntrdmax là khoảng cách tâm tối đa của các khung hình chữ nhật được giới hạn và Cntrdth là ngưỡng của khoảng cách trung tâm. Sự khác biệt giữa MulPer1th và MulPer2th là MulPer1th được sử dụng để đánh giá số lượng người khi chỉ có một khung hình chữ nhật có giới hạn, trong khi MulPer2th được sử dụng khi có nhiều hơn một khung hình chữ nhật có giới hạn.

Chúng tôi xác định trạng thái của hình ảnh như sau: (1) Khi có một khung hình chữ nhật được giới hạn trong hình ảnh,

a) Nếu Areamax < AreaThlow, thì hình ảnh này chỉ chứa một mục tiêu và được đánh dấu là Trạng thái 1;
b) Nếu AreaThlow ≤ Areamax ≤ AreaThup, số lượng mục tiêu trong hình ảnh này là không chắc chắn và

hình ảnh này được đánh dấu là Trạng thái 3; c) Nếu Areamax > AreaThup, thì hình ảnh này chứa nhiều mục tiêu và được đánh dấu là Trạng thái 4. (2) Khi có nhiều hơn một khung hình chữ nhật được giới hạn trong ảnh,

a) Nếu Cntrdmax > Cntrdth, thì các mục tiêu trong hình ảnh cách xa nhau và hình ảnh này được đánh dấu là Trạng thái 6; b) Nếu Cntrdmax ≤ Cntrdth và Areamax ≥ MulPer2th, thì tình hình trong hình ảnh này là không chắc chắn, và hình ảnh này được đánh dấu là Trạng thái 3; c) Nếu Cntrdmax ≤ Cntrdth, Areamin < AreaThlow và số lượng hình chữ nhật được giới hạn

khung hình là hai, thì khung hình nhỏ hơn không phải là con người và hình ảnh này được đánh dấu là Trạng thái 1; d) Nếu không có điều kiện nào ở trên được đáp ứng, thì hình ảnh được đánh dấu là Trạng thái 2. (3) Hơn nữa, xem xét trạng thái của khung hình (hình ảnh) trước đó,

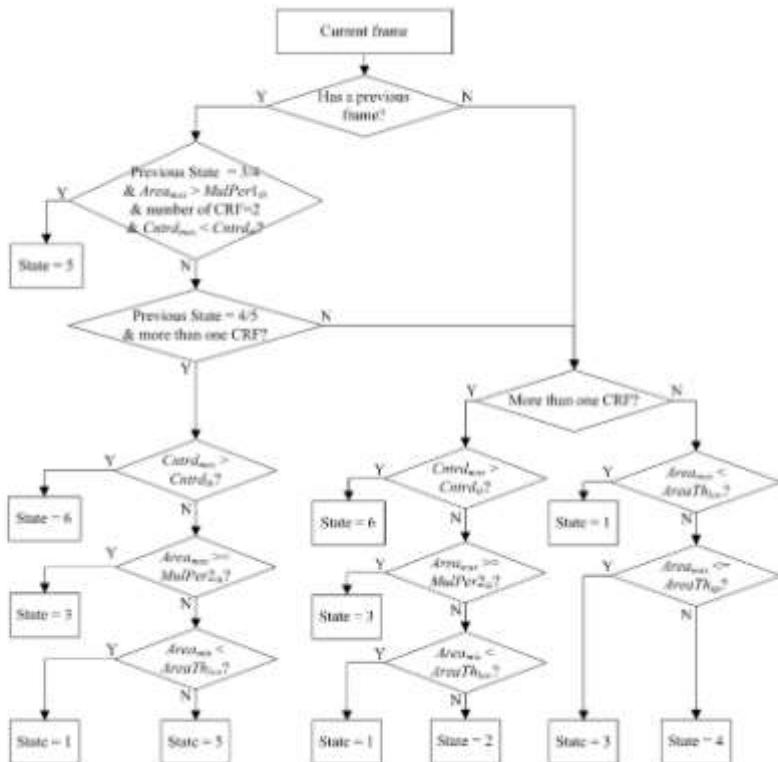
a) Nếu trạng thái của hình ảnh trước đó là Trạng thái 3 hoặc Trạng thái 4, Areamax > MulPer1th, số lượng mục tiêu trong hình ảnh là 2 và Cntrdmax < Cntrdth, điều này có thể có nghĩa là hai người đã gặp nhau nhưng nhanh chóng tách ra và hình ảnh này được đánh dấu là Trạng thái 5;
b) Nếu trạng thái của hình ảnh trước đó là Trạng thái 4 hoặc Trạng thái 5 và số lượng giới hạn

khung hình chữ nhật lớn hơn 1,

i) Nếu Cntrdmax > Cntrdth, thì hình ảnh được đánh dấu là Trạng thái 6;

- ii) Nếu $Cntrdmax \leq Cntrdth$ và $Areamax \geq MulPer2th$, trạng thái của các mục tiêu trong hình ảnh là không chắc chắn và hình ảnh được đánh dấu là Trạng thái 3; iii) Nếu $Cntrdmax \leq Cntrdth$, $Areamin < AreaThlow$ và số lượng giới hạn khung hình chữ nhật là hai, thì một trong các mục tiêu được phát hiện không phải là con người và hình ảnh được đánh dấu là Trạng thái 1; iv) Nếu không có điều kiện nào ở trên được đáp ứng, hình ảnh được đánh dấu là Trạng thái 5.

Hình 8 minh họa lưu đồ định nghĩa trạng thái. Các tính năng này rất trực quan để đánh giá chuyển động của các mục tiêu, vì vậy chúng tôi đã thiết kế một bộ phân loại Cây quyết định để chúng xác định một số hoạt động điển hình. Các chi tiết được mô tả trong Phần 5.



Hình 8. Lưu đồ định nghĩa trạng thái (CRF, khung hình chữ nhật giới hạn).

4.1.2. Tính năng dòng quang Luồng quang học là một phương pháp thường được sử dụng để phân tích mục tiêu di chuyển. Một đối tượng được biểu thị bằng pixel trong hình ảnh và khi đối tượng di chuyển, độ chói của các pixel sẽ thay đổi. Sự thay đổi của độ chói được gọi là dòng quang. Dòng quang học có thể phản ánh chuyển động của một vật thể. Có hai loại luồng quang chính: dòng quang dày đặc và dòng quang thưa. Dòng quang dày đặc thực hiện đăng ký hình ảnh với từng pixel trên hình ảnh và tính toán độ lệch cho tất cả các pixel, vì vậy nó có chi phí tính toán cao. Vì công việc này dành cho các ứng dụng thời gian thực nên dòng quang dày đặc không phù hợp. Mặt khác, luồng quang học thưa thực hiện đăng ký hình ảnh với các điểm thưa thớt trên hình ảnh. Cho một số điểm (thường là điểm góc) trên hình ảnh tham chiếu, thưa thớt

Luồng quang học tìm kiếm các điểm tương ứng trên hình ảnh mẫu. Giả sử rằng $P(x, y)$ là một điểm góc trên hình ảnh tham chiếu R và $Q(x + u, y + v)$ là điểm tương ứng trên hình ảnh mẫu S, trong đó u và v lần lượt là độ lệch trong trục x và trục y. Nếu tất cả các pixel trong một hộp hình chữ nhật nhỏ ở giữa P giống với các pixel trong cùng một hộp hình chữ nhật ở giữa Q, thì Q khớp với P. Tuy nhiên, có nhiều trong hình ảnh thực tế, vì vậy các pixel trong hai hộp không thể giống nhau. Trong trường hợp này, lấy điểm có chênh lệch nhỏ nhất làm điểm tham chiếu:

$$\begin{aligned} P(x, y) = \operatorname{argmin}_{u, v} E(u, v) &= \sum_{(x, y) \in W} \|R(x, y) - S(x + u, y + v)\| \\ \end{aligned} \quad (3)$$

trong đó W là hộp hình chữ nhật ở giữa P. Phương trình (3) có thể được giải bằng một kim tự tháp hình ảnh. Chúng tôi trích xuất cực đại, cực đại thứ 10 [23] và trung bình của biên độ dòng quang làm các tính năng.

4.2. Chuẩn hóa điểm Z
Nếu một mô hình phân loại không có đặc tính bất biến tỷ lệ, các tính năng có độ lớn khác nhau sẽ khiến các tham số mô hình bị chi phối bởi dữ liệu lớn hơn hoặc nhỏ hơn, và do đó quy trình đào tạo sẽ bị ảnh hưởng tiêu cực. Khi so sánh sự đóng góp của các tính năng có cường độ khác nhau đối với phân loại, chuẩn hóa thường là cần thiết. Trong công việc này, chúng tôi sử dụng thuật toán chuẩn hóa Z-Score, được tính như

$$x' = x - \mu\sigma \quad (4)$$

Chuẩn hóa Z-Score rất đơn giản và do đó có chi phí tính toán thấp. Sau khi chuẩn hóa Z-Score, quy trình đào tạo có thể tránh được vấn đề không ổn định trong tính toán số do trọng lượng không cân bằng.

4.3. Lựa chọn tính năng
Vì không phải tất cả các tính năng được trích xuất đều có thể góp phần vào việc phân loại, nên việc lựa chọn tính năng thường là một thủ tục cần thiết trong nhận dạng mẫu. Chúng tôi đã áp dụng các thuật toán lựa chọn tính năng khác nhau trong các giai đoạn nhận dạng hoạt động không quan tâm.

4.3.1. Relief-F
Relief-F được sử dụng ngay sau khi các tính năng hình ảnh được trích xuất. Relief-F là một thuật toán kiểu Bộ lọc được phát triển từ Relief. Nó xem xét k mẫu gần nhất thay vì chỉ một mẫu và hỗ trợ phân loại đa lớp. Relief-F cung cấp các tính năng có trọng số khác nhau dựa trên mức độ liên quan của từng tính năng và danh mục. Mỗi tương quan giữa các tính năng và danh mục trong Thuật toán cứu trợ dựa trên khả năng phân biệt giữa các mẫu tầm gần của các tính năng. Các tính năng có trọng số nhỏ hơn một ngưỡng nhất định sẽ bị xóa. Trong bài báo này, chúng tôi đã sử dụng Relief-F proposed cải tiến trong công việc trước của chúng tôi [4], giúp giảm thêm tính năng dư thừa. Chúng tôi đã áp dụng Relief-F cho các tính năng khung hình chữ nhật (và mục tiêu được phát hiện), cũng như các tính năng luồng quang học được trích xuất trong Phần 3 và loại bỏ các tính năng có trọng số không dương. Cuối cùng, chúng tôi thu được 14 tính năng hữu ích: chiều rộng tối đa, biến đổi chiều rộng tối đa, chiều cao tối đa, thay đổi chiều cao tối đa, diện tích tối đa, biến đổi diện tích tối đa, tỷ lệ khung hình tối đa, khoảng cách tâm tối đa, biến đổi khoảng cách tâm tối đa của khung hình chữ nhật giới hạn, tổng diện tích, diện tích tối đa, trạng thái của các mục tiêu được phát hiện và trung bình của luồng quang. Sau đó, có các tính năng được sử dụng để chọn hạt nhân tốt nhất cho bộ phân loại SVM.

4.3.2. Wrapper được sử dụng sau khi chúng ta xác định hàm hạt nhân của SVM vì Wrapper cần bộ phân loại để đánh giá sự đóng góp của các tính năng. Wrapper có thể được thực hiện tiến hoặc lùi hoặc theo các kết hợp khác (ví dụ: tiến-lùi và lùi-tiến). Vì đã có 14 tính năng trong bộ tính năng, chúng tôi đã sử dụng Wrapper lùi-tiến. Để rõ ràng, chúng tôi đánh số các tính năng như trong Bảng 1.

Bảng 1. Đánh số các tính năng.

| Tính năng | Số |
|--|------|
| Chiều rộng tối đa của (các) CRF 1 | (1) |
| Biến đổi chiều rộng tối đa của (các) CRF | (2) |
| Chiều cao tối đa của (các) CRF | (3) |
| Biến thể chiều cao tối đa của (các) CRF | (4) |
| Điện tích tối đa của (các) CRF | (5) |
| Sự thay đổi điện tích tối đa của (các) CRF | (6) |
| Tỷ lệ khung hình tối đa của (các) CRF | (7) |
| Biến thể tỷ lệ khung hình tối đa của (các) CRF | (8) |
| Khoảng cách tâm tối đa của (các) CRF | (9) |
| Sự thay đổi khoảng cách tâm tối đa của (các) CRF | (10) |
| Tổng diện tích các mục tiêu được phát hiện | (11) |
| Điện tích tối đa của mục tiêu được phát hiện | (12) |
| Trạng thái mục tiêu được phát hiện | (13) |
| Trung bình của dòng quang | (14) |

1 (các) CRF: (các) khung hình chữ nhật có giới hạn.

Đầu tiên, một quy trình "ngược" đã được thực hiện. Chúng tôi đã xóa từng tính năng khỏi bộ tính năng và quan sát độ chính xác nhảm dặng, bắt đầu bằng Tính năng (1). Bảng 2 cho thấy sự thay đổi độ chính xác trong quá trình "ngược".

Bảng 2. Sự thay đổi độ chính xác trong quá trình "lùi".

| Đã xóa (các) tính năng | Độ chính xác (%) | Hành động |
|---|------------------|--------------|
| Không ai | 89,15 | - |
| (1) | 89,02 ↓ | Giữ lại (1) |
| (2) | 89,03 ↓ | Giữ lại (2) |
| (3) | 88,82 ↓ | Giữ lại (3) |
| (4) | 89,05 ↓ | Giữ lại (4) |
| (5) | 89,12 ↓ | Giữ lại (5) |
| (6) | 89,12 ↓ | Giữ lại (6) |
| (7) | 88,60 ↓ | Giữ lại (7) |
| (8) | 89,22 ↑ | Loại bỏ (8) |
| (8), (9)89,42 ↑Xóa (9)(8), (9), (10)89,35 ↓Giữ lại (10)(8), (9), (11)88,94 ↓Giữ lại (11)(8), (9), (12)89,42 =Thứ (12) sau(8), (9), (12), (13)87,72 ↓Giữ lại (13)(8), (9), (12), (14)89,62 ↑Xóa (14) | | |
| (8), (9), (14) | 89,52 ↓ | Loại bỏ (12) |

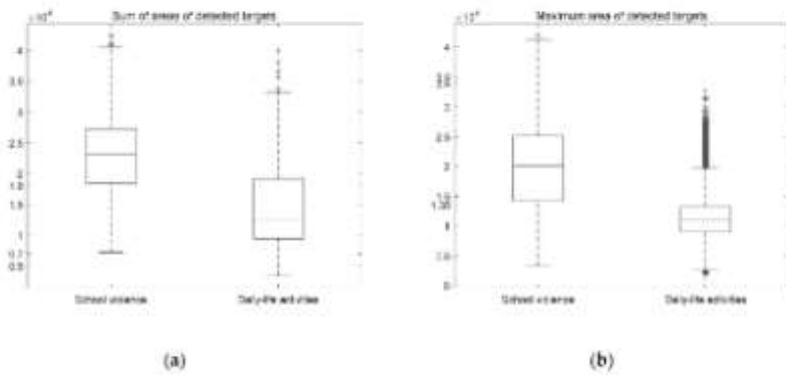
Sau quy trình "ngược", bốn đặc điểm đã được loại bỏ, đó là (8), (9), (12) và (14). Vì sự kết hợp các tính năng khác nhau có thể có các hiệu ứng khác nhau, để tránh loại bỏ các tính năng do nhầm lẫn, một quy trình "chuyển tiếp" đã được thực hiện. Bảng 3 cho thấy sự thay đổi độ chính xác trong quy trình "chuyển tiếp".

Bảng 3. Sự thay đổi độ chính xác trong quá trình "chuyển tiếp".

| (Các) tính năng được thêm vào | Độ chính xác (%) | Hành động |
|-------------------------------|------------------|--------------|
| Không ai | 89,62 | - |
| (8) | 89,48 ↓ | Loại bỏ (8) |
| (9) | 88,86 ↓ | Loại bỏ (9) |
| (12) | 89,52 ↓ | Loại bỏ (12) |
| (14) | 89,42 ↓ | Loại bỏ (14) |

Trong quy trình "chuyển tiếp", không có tính năng nào trong số bốn tính năng bị loại bỏ có thể cải thiện độ chính xác, vì vậy bốn tính năng này đã bị loại bỏ khỏi bộ tính năng. Sau khi chọn tính năng Wrapper, độ chính xác tăng từ 89,15% lên 89,62%. Vì kích thước tính năng sau khi chọn tính năng không cao, chúng tôi không thực hiện thêm các quy trình giảm kích thước. Sau khi lựa chọn tính năng Relief-F và Wrapper, tất cả các tính năng còn lại đều đóng góp tích cực vào việc phân loại; Giảm kích thước hơn nữa sẽ làm mất thông tin hữu ích và do đó làm giảm độ chính xác của nhận dạng.

4.3.3. Boxplot Một boxplot được sử dụng để tìm các tính năng thích hợp và ngưỡng thích hợp để thiết kế DecisionTree. Như đã đề cập trong Phần 3, một số đặc điểm có thể phân biệt các hoạt động diễn hình với các hoạt động khác. Hình 8 cho thấy hai ví dụ về các đặc điểm như vậy, cụ thể là tổng diện tích của các mục tiêu được phát hiện và diện tích tối đa của các mục tiêu được phát hiện. Hình 9a cho thấy sự phân bố tổng các khu vực của các mục tiêu được phát hiện về bạo lực học đường và các hoạt động sinh hoạt hàng ngày. Giá trị $1,8 \times 104$ là ngưỡng ngăn cách 75% bạo lực học đường với 75% các hoạt động hàng ngày. Tuy nhiên, ngưỡng này tốt cho SVM nhưng không hoàn hảo cho DT. Giá trị $3,4 \times 104$ là một ngưỡng khác. Không có phân phối mẫu hoạt động sinh hoạt hàng ngày nào trên ngưỡng này, vì vậy có thể tách một số mẫu bạo lực học đường diễn hình khỏi các hoạt động sinh hoạt hàng ngày.



Hình 9. Hai ví dụ về biểu đồ hộp: (a) tổng diện tích của các mục tiêu được phát hiện và (b) diện tích tối đa của các mục tiêu được phát hiện.

5. Thuật toán phát hiện bạo lực học đường DT-SVM

5.1. Phân loại SVM Vì mục đích của công việc này là phân loại 2 lớp (tức là bắt nạt thể chất và các hoạt động sinh hoạt hàng ngày), chúng tôi đã chọn SVM làm bộ phân loại. SVM đã được chứng minh là một lựa chọn tốt để phân loại 2 lớp. Hơn nữa, mục đích của công việc này là thiết kế một hệ thống phát hiện bạo lực theo thời gian thực, vì vậy chi phí tính toán của thuật toán phát hiện phải càng thấp càng tốt. So với các phương pháp học máy khác, chẳng hạn như ANN (Mạng neural) và các phương pháp học sâu, SVM có chi phí tính toán thấp hơn. SVM cung cấp một số hàm hạt nhân để phù hợp với các phân phối dữ liệu khác nhau, cụ thể là hàm hạt nhân tuyến tính, hàm nhân đa thức, hàm cơ sở xuyên tâm (RBF) và hàm hạt nhân sigmoid. Chúng tôi đã kiểm tra hiệu suất phân loại của bốn hàm hạt nhân. Đầu vào là 14 tính năng được chọn bởi Relief-F sau khi chuẩn hóa Z-Score. Xác nhận chéo năm lần đã được sử dụng và kết quả mô phỏng được đưa ra trong Bảng 4.

Bảng 4. So sánh bốn chức năng hạt nhân.

| Chức năng hạt nhân | Độ chính xác (%) |
|--------------------|------------------|
| Tuyến Đa thức | 86.7 |
| RBF | 89.2 |
| Sigmoid | 78.1 |

Có thể thấy từ Bảng 4 rằng RBF cung cấp hiệu suất tốt nhất, vì vậy chúng tôi đã chọn RBF làm hàm hạt nhân trong công việc sau.

5.2. Phân loại DT-SVM Như đã đề cập trong Phần 4, một số tính năng có thể phân biệt chính xác các hoạt động khác nhau với nhau, vì vậy chúng tôi đã thiết kế Cây quyết định cho các tính năng này. So với các bộ phân loại khác, DecisionTree có một số ưu điểm, chẳng hạn như chi phí tính toán thấp và khả năng chống nhiễu. Nhiệm vụ chính trong việc thiết kế Cây quyết định là tìm ra các tính năng thích hợp và ngưỡng thích hợp của chúng. Hình 8 trong Phần 4 cho thấy hai ví dụ về cách tìm các đối tượng địa lý thích hợp và ngưỡng tương ứng của chúng. Chúng tôi lặp lại quy trình tương tự cho các tính năng và ngưỡng có thể có khác. Bảng 5 hiển thị các đối tượng địa lý đã chọn và các ngưỡng tương ứng của chúng.

Bảng 5. Các tính năng được chọn cho DT (Cây quyết định) và các ngưỡng tương ứng của chúng (đơn vị: pixel2 for area và pixel cho các tính năng khác).

| Tính năng | Nguồng | Phân loại là |
|--|----------|-------------------------------|
| Diện tích tối đa của mục tiêu được phát hiện | > 40.000 | Bạo lực học đường |
| Biến đổi diện tích tối đa của (các) CRF 1 | > 50.000 | Bạo lực học đường |
| Tổng diện tích các mục tiêu được phát hiện | < 6000 | Hoạt động sinh hoạt hàng ngày |
| Số lượng CRF | 0 | Hoạt động sinh hoạt hàng ngày |
| Khoảng cách tâm tối đa của (các) CRF | > 300 | Hoạt động sinh hoạt hàng ngày |
| Chiều rộng tối đa của (các) CRF | < 50 | Hoạt động sinh hoạt hàng ngày |
| Diện tích rộng tối đa của (các) CRF | < 10.000 | Hoạt động sinh hoạt hàng ngày |
| Sự thay đổi khoảng cách tâm tối đa của (các) CRF | > 400 | Hoạt động sinh hoạt hàng ngày |
| Trạng thái của các mục tiêu | = 1 | Hoạt động sinh hoạt hàng ngày |
| Trung bình của dòng quang | > 2000 | Hoạt động sinh hoạt hàng ngày |

1 (các) CRF: (các) khung hình chữ nhật có giới hạn.

Để chung chung hơn, các ngưỡng đã được đưa ra một biên độ lỗi. Ví dụ, trong Hình 9a, ngưỡng chính xác để tách các hoạt động sinh hoạt hàng ngày điển hình khỏi bạo lực học đường là 7000 pixel2,

nhưng nói chung hơn, chúng tôi đặt ngưỡng trong DT là 6000 pixel². Điều tương tự cũng được thực hiện đối với các ngưỡng khác.

DT và SVM hoạt động ở chế độ 2 lớp. Nếu DT có thể xác định chính xác hoạt động được thử nghiệm (nếu một trong các điều kiện trong Bảng 5 được đáp ứng), thì DT xác định kết quả phân loại; nếu không, SVM sẽ xác định kết quả phân loại.

6. Kết quả thí nghiệm

Trong thí nghiệm này, chúng tôi đã chụp được tổng cộng 24.896 khung hình hoạt động, bao gồm 12.448 khung hình bạo lực học đường, 9963 khung hình hoạt động cuộc sống hàng ngày và 2485 khung hình tĩnh. Sau khi lựa chọn tính năng, 10 tính năng đã được sử dụng để phân loại SVM: chiều rộng tối đa, biến đổi chiều rộng tối đa, chiều cao tối đa, biến đổi chiều cao tối đa, diện tích tối đa, biến đổi diện tích tối đa, tỷ lệ khung hình tối đa, biến thể khoảng cách tâm tối đa của khung hình chữ nhật giới hạn và tổng diện tích và trạng thái của các mục tiêu được phát hiện. Các tính năng này và các ngưỡng tương ứng được sử dụng cho DT được đưa ra trong Mục 5.2.

Chúng tôi định nghĩa bạo lực học đường là tích cực và hoạt động cuộc sống hàng ngày là tiêu cực, vì vậy TP (True Positive) có nghĩa là bạo lực học đường được công nhận là bạo lực học đường, FP (False Positive) có nghĩa là hoạt động cuộc sống hàng ngày được công nhận là bạo lực học đường (còn được gọi là báo động giả), TN (True Negative) có nghĩa là hoạt động cuộc sống hàng ngày được công nhận là hoạt động cuộc sống hàng ngày và FN (False Negative) có nghĩa là bạo lực học đường được công nhận là hoạt động cuộc sống hàng ngày (còn được gọi là báo động thiếu). Bốn chỉ số sau đây được sử dụng để đánh giá hiệu suất phân loại:

$$\text{độ chính xác} = \frac{|TP|}{|TN||P| + |N|}, \quad (5)$$

$$\text{Độ chính xác} = \frac{|TP||TP|}{|FP|}, \quad (6)$$

$$\text{Nhớ lại} = \frac{|TP||TP|}{|FN|}, \quad (7)$$

$$F1 - Điểm = 2 \times \frac{\text{độ chính xác}}{\text{độ chính xác} + \text{thu hồi}} \quad (8)$$

Thứ nhất, chỉ có SVM được sử dụng để phân loại. RBF được sử dụng làm hàm hạt nhân, và xác nhận năm lần chéo đã được sử dụng. Bảng 6 cho thấy ma trận nhầm lẫn của phân loại SVM. Độ chính xác = 89,6%, độ chính xác = 94,4%, thu hồi = 81,5% và Điểm F1 = 87,5%.

Bảng 6. Ma trận nhầm lẫn của phân loại SVM (Máy vectơ hỗ trợ).

| Phân loại là | Bạo lực học đường | Hoạt động hàng ngày |
|-------------------------------|-------------------|---------------------|
| Bạo lực học đường | 96,1% | 3,9% |
| Hoạt động sinh hoạt hàng ngày | 18,5% | 81,5% |

Sau đó, chúng tôi sử dụng DT-SVM để phân loại. Bảng 7 cho thấy ma trận nhầm lẫn của phân loại DT-SVM.

Bảng 7. Ma trận nhầm lẫn của phân loại DT-SVM (Cây quyết định – Máy vectơ hỗ trợ).

| Phân loại là | Bạo lực học đường | Hoạt động hàng ngày |
|-------------------------------|-------------------|---------------------|
| Bạo lực học đường | 96,8% | 3,2% |
| Hoạt động sinh hoạt hàng ngày | 1,7% | 98,3% |

Độ chính xác = 97,6%, độ chính xác = 97,2%, thu hồi = 98,3% và Điểm F1 = 97,8%. Có thể thấy rằng Cây quyết định được thiết kế cải thiện đáng kể độ chính xác nhận dạng của các hoạt động trong cuộc sống hàng ngày, từ

81,5% đến 98,3%. Thuật toán phát hiện bạo lực học đường được đề xuất cũng có hiệu suất tốt hơn so với các thuật toán phát hiện bạo lực hiện có (ví dụ: 96,50% trong [16] và 91% trong [18]).

7. Thảo luận và kết luận

Bạo lực học đường là một vấn đề xã hội phổ biến gây hại cho thanh thiếu niên. May mắn thay, hiện nay có một số phương pháp có thể phát hiện bạo lực học đường, chẳng hạn như phương pháp sử dụng cảm biến chuyển động và phương pháp sử dụng camera. Chúng tôi đã sử dụng cảm biến chuyển động để phát hiện bạo lực học đường trong nghiên cứu trước đây của mình, nhưng trong bài báo này, chúng tôi đã chọn một phương pháp khác - một phương pháp sử dụng camera trong trường hợp cảm biến chuyển động bị loại bỏ bởi những kẻ bắt nạt. Máy ảnh được sử dụng để chụp ảnh khu vực giám sát và phát hiện các mục tiêu di chuyển. Chúng tôi đã đề xuất một phương pháp tích hợp khung hình chữ nhật giới hạn để tối ưu hóa mục tiêu tiềm cảnh được phát hiện. Sau đó, các đặc điểm khung hình chữ nhật và các đặc điểm dòng chảy quang học được trích xuất để mô tả sự khác biệt giữa bạo lực học đường và các hoạt động trong cuộc sống hàng ngày. Relief-F và Wrapperđược sử dụng để giảm kích thước tính năng. Sau đó, một bộ phân loại DT-SVM đã được xây dựng để phân loại. Độ chính xác đạt 97,6% và độ chính xác đạt 97,2%. Bằng cách phân tích kết quả mô phỏng, chúng tôi xác định rằng hình ảnh có sự thay đổi lớn về ánh sáng và bóng tối và các hành động bạo lực với biên độ nhẹ dễ dàng bị phân loại sai. Công việc này cho thấy hứa hẹn cho việc giám sát bạo lực trong khuôn viên trường. Trong tương lai, chúng tôi dự định liên quan đến các hoạt động phức tạp hơn, chẳng hạn như bắt nạt / đánh nhau theo nhóm và các cảnh phức tạp hơn với cả các đối tượng gần và xa, đồng thời cải thiện hiệu suất nhận dạng của các mẫu được phân loại theo chủ đề.

Đóng góp của tác giả: Khái niệm, LY; phương pháp luận, L.Y. và L.W.; phần mềm, LW và LY; xác nhận, LW và LY; phân tích chính thức, LW và LY; điều tra, L.Y. và L.W.; quản lý dữ liệu, LY; viết—chuẩn bị bản thảo gốc, L.Y. và L.W.; viết—dánh giá và chỉnh sửa, H.F., T.S., và E.A.; hình dung, LY và LW; giám sát, LY, EA và TS; quản trị dự án, EA và LY; mua lại tài trợ, L.Y. và H.F. Allauthors đã đọc và đồng ý với phiên bản xuất bản của bản thảo. Tài trợ: Nghiên cứu này được tài trợ bởi Quỹ Khoa học Tự nhiên Quốc gia Trung Quốc, số tài trợ 41861134010; Đề tài nghiên cứu khoa học cơ bản của tỉnh Hồ Bắc Long Giang, số tài trợ KJCXZD201704; Phòng thí nghiệm trọng điểm của Cảnh sát Truyền thông kỹ thuật số không dây, Bộ Công an, số tài trợ 2018JYWXTX01, và một phần do Quỹ Văn hóa Phần Lan, Quỹ Khu vực Bắc Ostrobothnia 2017.**Lời cảm ơn:** Các tác giả muốn cảm ơn những người đã giúp đỡ những thí nghiệm này. Xung đột lợi ích: Các tác giả tuyên bố không có xung đột lợi ích.

Tham khảo

- 1.Ye, L.; Ferdinando, H.; Seppänen, T.; Alasaarela, E. Phát hiện bạo lực học đường bằng cảm biến bắt nạt học đường. *Trí tuệ*. 2014, 2014, 740358. [Tham khảo chéo]
- 2.Ye, L.; Ferdinando, H.; Seppänen, T.; Huuki, T.; Alasaarela, E. Thuật toán phát hiện bạo lực học đường dựa trên cảm biến bắt nạt học đường. Trong *Ký yếu của Hội nghị Truyền thông Không dây và Điện toán Di động Quốc tế (IWCMC)* năm 2015, Dubrovnik, Croatia, 24–28 tháng 8 năm 2015; trang 1384–1388.
- 3.Ye, L.; Vương, P.; Vương, L.; Ferdinando, H.; Seppänen, T.; Alasaarela, E. Một thuật toán phát hiện bắt nạt học đường kết hợp giữa chuyển động và âm thanh. *Int. J. Nhận dạng mẫu*. Nghệ thuật. *Trí tuệ*. 2018, 32, 1850046. [Tham khảo chéo]
- 4.Ye, L.; Shi, J.; Ferdinando, H.; Seppänen, T.; Alasaarela, E. Phát hiện bạo lực học đường dựa trên đa giác quan và các thuật toán Relief-F được cải thiện. Trong *Ký yếu của Hội nghị Quốc tế về Trí tuệ nhân tạo cho Truyền thông và Mạng 2019*, Cáp Nhĩ Tân, Trung Quốc, 25–26 tháng 5 năm 2019.
- 5.Zainudin, MS; Sulaiman, M.N.; Mustapha, N.; Perumal, T. Nhận dạng hoạt động bằng chiến lược một so với tất cả với Relief-F và thuật toán tự thích ứng. Trong *Ký yếu của Hội nghị IEEE 2018 về Hệ thống Mở (ICOS)*, Đảo Langkawi, Malaysia, 21–22 tháng 11 năm 2018.
- 6.Bhavan, A.; Aggarwal, S. Tổng quát hóa xếp chồng với lựa chọn tính năng dựa trên trình bao bọc để nhận dạng hoạt động của con người. Trong *Ký yếu của Chuỗi Hội nghị chuyên đề IEEE 2018 về Trí tuệ tính toán (SSCI)*, Bangalore, Ấn Độ, 18–21 tháng 11 năm 2018.
- 7.Mohammed, N.N.; Khaleel, M.; Latif, M.; Khalid, Z. Nhận dạng khuôn mặt dựa trên PCA với khoảng cách Mahalanobis có trọng số và chuẩn hóa. Trong *Ký yếu của Hội nghị Quốc tế 2018 về Tin học thông minh và Khoa học Y sinh (ICIBMS)*, Bangkok, Thái Lan, 21–24 tháng 10 năm 2018.

8. Aburomman, A.A.; Reaz, MBI. Tập hợp các bộ phân loại SVM nhị phân dựa trên trích xuất tính năng PCA và LDA để phát hiện xâm nhập. Trong Ký yếu của Hội nghị Quản lý Thông tin Nâng cao, Truyền thông, Điều khiển Điện tử và Tự động hóa (IMCEC) năm 2016, Tây An, Trung Quốc, ngày 3–5 tháng 10 năm 2016.
9. Hán, S.; Trương, Y.; Meng, W.; Li, C.; Zhang, Z. Macrocell hỗ trợ role song công với sóng truyền millimet: Khâu khở và triển vọng. *IEEE Netw.* 2019, **33**, 190–197. [Tham khảo chéo]
10. Attal, F.; Mohammed, S.; Dedabirashvili, M.; Chamroukhi, F.; Oukhellou, L.; Amarat, Y. Nhận dạng hoạt động thể chất của con người bằng cách sử dụng cảm biến có thể đeo được. *Cảm biến* 2015, **15**, 31314–31338. [Tham khảo chéo] [PubMed]
11. Lu, W.; Gong, Y.; Lưu, X.; Ngô, J.; Peng, H. Truyền thông tin và năng lượng hợp tác trong mạng cảm biến không dây xanh cho thành phố thông minh. *IEEE Trans. Ind. Thông báo.* 2017, **14**, 1585–1593. [Tham khảo chéo]
12. Lưu, X.; Gia Á, M.; Na, Z. Cảm biến phổ hợp tác đa phương thức dựa trên sự hợp nhất dempster-shafer trong radio nhận thức dựa trên 5G. *Truy cập IEEE* 2018, **6**, 199–208. [Tham khảo chéo]
13. Mặt trời, X.; Huang, D.; Vương, Y.; Qin, J. Nhận dạng hành động dựa trên biểu diễn động học của dữ liệu video. Trong Ký yếu của Hội nghị Quốc tế IEEE 2014 về Xử lý Hình ảnh (ICIP), Paris, Pháp, 27–30 tháng 10 năm 2014.
14. Vương, H.; Nguyễn thủy, C.; Hu, W.; Ling, H.; Dương, W.; Sun, C. Nhận dạng hành động bằng cách sử dụng biểu diễn thành phần hành động không tiêu cực và lựa chọn cơ sở thưa thớt. *IEEE Trans. Quy trình hình ảnh.* 2014, **23**, 570–581. [Tham khảo chéo] [PubMed]
15. Tu, Z.; Li, H.; Trương, D.; Dauwels, J.; Li, B.; Yuan, J. Giai đoạn hành động nhấn mạnh VLAD không gian thời gian để nhận dạng hành động video. *IEEE Trans. Quy trình hình ảnh.* 2019, **28**, 2799–2812. [Tham khảo chéo] [PubMed]
16. Keçeli, AS; Kaya, A. Phát hiện hoạt động bạo lực bằng phương pháp học chuyển giao. *Điện tử. Lett.* 2017, **53**, 1047–1048. [Tham khảo chéo]
17. Ha, J.; Công viên, J.; Kim, H.; Công viên, H.; Paik, J. Phát hiện bạo lực cho hệ thống giám sát video bằng cách sử dụng thông tin chuyển động bất thường. Trong Ký yếu của Hội nghị Quốc tế về Điện tử, Thông tin và Truyền thông (ICEIC) năm 2018, Honolulu, HI, Hoa Kỳ, ngày 24–27 tháng 1 năm 2018.
18. Ehsan, T.Z.; Nahvi, M. Phát hiện bạo lực trong camera giám sát trong nhà bằng cách sử dụng quỹ đạo chuyển động và biểu đồ vi sai của luồng quang học. Trong Ký yếu của Hội nghị Quốc tế lần thứ 8 năm 2018 về Kỹ thuật Máy tính và Tri thức (ICCKE), Mashhad, Iran, 25–26 tháng 10 năm 2018; trang 153–158.
19. Shu, Z.; Nannan, C.; Chao, F.; Thiếu Sinh, L. Việc thiết kế và triển khai hệ thống servo camera tự động dựa trên thuật toán vi sai khung hình được cải thiện. Trong Ký yếu của Hội nghị Quốc tế lần thứ 2 năm 2011 về Trí tuệ nhân tạo, Khoa học Quản lý và Thương mại Điện tử (AIMSEC), Dengleng, Trung Quốc, ngày 8–10 tháng 8 năm 2011.
20. Tian, Y.; Đăng, L.; Lý, Q. Phương pháp theo dõi-học tập-phát hiện dựa trên KNN phù hợp với việc điều chỉnh các khu vực được khảo sát. Trong Ký yếu của Hội nghị Quốc tế lần thứ 13 năm 2017 về Trí tuệ tính toán và An ninh (CIS), Hồng Kông, Trung Quốc, 15–18 tháng 12 năm 2017.
21. Benraya, I.; Benblidia, N. So sánh các phương pháp trừ nền. Trong Ký yếu của Hội nghị Quốc tế về Hệ thống Thông minh Ứng dụng (ICASS) năm 2018, Medéa, Algeria, ngày 24–25 tháng 11 năm 2018.
22. Cao Đông, W.; Trương, X.; Dương, L.; Lưu, H. Phát hiện cạnh sobel được cải thiện. Trong Ký yếu của hội nghị quốc tế lần thứ 3 năm 2010 về khoa học máy tính và công nghệ thông tin, Thành Đô, Trung Quốc, ngày 9–11 tháng 7 năm 2010.
23. Ji, Y.; Trần, Z.; Ma, Z.; Wang, F. Phát hiện bạo lực hành khách dựa trên video thang máy. *Ind. Kiểm soát Comput.* 2018, **31**, 1–3.



© 2020 bởi các tác giả. Bản được cấp phép MDPI, Basel, Thụy Sĩ. Bài viết này là truy cập mở được phân phối theo các điều khoản và điều kiện của giấy phép Creative Commons Chi công (CC BY) (<http://creativecommons.org/licenses/by/4.0/>).