Biểu đồ của Gradient định hướng để phát hiện con người

Navneet Dalal và Bill Triggs
INRIA Rhône-Alps, 655 Avenue de l'Europe, Montbonnot 38334, Pháp
{Navneet.Dalal,Bill.Triggs}@inrialpes.fr, http://lear.inrialpes.fr

trừu tượng

Chúng tôi nghiên cứu câu hỏi về bộ tính năng để nhận dạng đối tượng trực quan mạnh mẽ, áp dụng phát hiện con người dựa trên SVM tuyến tính làm trường hợp thử nghiệm. Sau khi xem xét các bộ mô tả dựa trên cạnh và độ dốc hiện có, chúng tôi cho thấy bằng thực nghiệm rằng các lưới của các bộ mô tả Biểu đồ độ dốc định hướng (HOG) hoạt động tốt hơn đáng kể so với các bộ tính năng hiện có để phát hiện con người. Chúng tôi nghiên cứu ảnh hưởng của từng giai đoạn tính toán đối với hiệu suất, kết luận rằng độ dốc tỷ lệ nhỏ, tạo khung định hướng tốt, tạo khung không gian tương đối thô và chuẩn hóa độ tương phản cục bộ chất lượng cao trong các khối mô tả chồng chéo đều quan trọng để mang lại kết quả tốt. Phương pháp mới mang lại sự phân tách gần như hoàn hảo trên cơ sở dữ liệu ban đầu của MIT dành cho người đi bộ,

1. Giới thiệu

Phát hiện con người trong hình ảnh là một nhiệm vụ đầy thách thức do ngoại hình thay đổi của họ và nhiều tư thế mà họ có thể áp dụng. Nhu cầu đầu tiên là một bộ tính năng mạnh mẽ cho phép phân biệt rõ ràng hình dạng con người, ngay cả trong nền lộn xộn dưới ánh sáng khó. Chúng tôi nghiên cứu vấn đề về các bộ tính năng để phát hiện con người, cho thấy rằng các bộ mô tả Biểu đồ độ dốc định hướng (HOG) được chuẩn hóa cục bộ cung cấp hiệu suất tuyệt vời so với các bộ tính năng hiện có khác bao gồm cả wavelet [17, 22]. Các bộ mô tả được đề xuất gợi nhớ đến biểu đồ định hướng cạnh [4, 5], bộ mô tả SIFT [12] và ngữ cảnh hình dạng [1], nhưng chúng được tính toán trên một lưới dày đặc các ô cách đều nhau và chúng sử dụng các chuẩn hóa tương phản cục bộ chồng lấp để cải thiện hiệu suất. Chúng tôi thực hiện một nghiên cứu chi tiết về tác động của các lựa chọn triển khai khác nhau đối với hiệu suất của máy dò, lấy "phát hiện người đi bộ" (phát hiện hầu hết những người có thể nhìn thấy ở các tư thế ít nhiều thẳng đứng) làm trường hợp thử nghiệm. Để đơn giản và nhanh chóng, chúng tôi sử dụng SVM tuyến tính làm công cụ phân loại cơ bản trong suốt quá trình nghiên cứu. Các công cụ phát hiện mới về cơ bản cho kết quả hoàn hảo trên bộ thử nghiệm dành cho người đi bộ của MIT [18, 17], vì vậy chúng tôi đã tạo ra một bộ thử thách hơn chứa hơn 1800 hình ảnh về người đi bộ với nhiều tư thế và bối cảnh khác nhau. Công việc đang tiến hành cho thấy rằng bộ tính năng của chúng tôi hoạt động tốt như nhau đối với các lớp đối tượng dựa trên hình dạng khác. chúng tôi sử dụng SVM tuyến tính làm công cụ phân loại cơ bản trong suốt quá trình nghiên cứu. Các công cụ phát hiện mới về cơ bản cho kết quả hoàn hảo trên bộ thử nghiệm dành cho người đi bộ của MIT [18, 17], vì vậy chúng tôi đã tạo ra một bộ thử thách hơn chứa hơn 1800 hình ảnh về người đi bộ với nhiều tư thế và bối cảnh khác nhau. Công việc đang tiến hành cho thấy rằng bộ tính năng của chúng tôi hoạt động tốt như nhau đối với các lớp đối tượng dựa trên hình dạng khác. chúng tôi sử dụng SVM tuyến tính làm công cụ phân loại cơ bản trong suốt quá trình nghiên cứu. Các công cụ phát hiện mới về cơ bản cho kết quả hoàn hảo trên bộ thử nghiệm dành cho người đi bộ của MIT [18, 17], vì vậy chúng tôi đã tạo ra một bộ thử thách hơn chứa hơn 1800 hình ảnh về người đi bộ với nhiều tư thế và bối cảnh khác nhau. Công việc đang tiến hành cho thấy rằng bộ tính năng của chúng tôi hoạt động tốt như nhau đối với các lớp đối tượng dựa trên hình dạng khác.

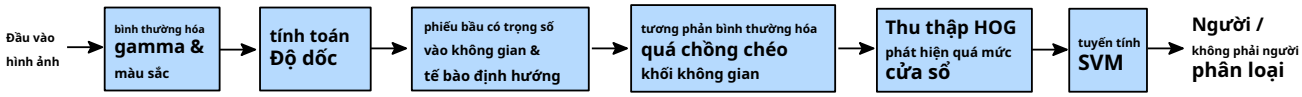
Chúng tôi thảo luận ngắn gọn về công việc trước đây về phát hiện con người trong §2, đưa ra một cái nhìn tổng quan về phương pháp của chúng tôi§3, mô tả bộ dữ liệu của chúng tôi trong§4 và đưa ra mô tả chi tiết và đánh giá thực nghiệm từng giai đoạn của quy trình trong§5–6. Các kết luận chính được tóm tắt trong§7.

2 công việc trước

Có nhiều tài liệu về phát hiện đối tượng, nhưng ở đây chúng tôi chỉ đề cập đến một số bài báo liên quan về phát hiện con người [18, 17, 22, 16, 20]. Xem [6] cho một cuộc khảo sát. Trang chủ*et al*[18] mô tả bộ phát hiện người đi bộ dựa trên SVM đa thức sử dụng sóng con Haar đã chỉnh lưu làm bộ mô tả đầu vào, với một biến thể dựa trên bộ phận (cửa sổ con) trong [17]. người nghèo*et al* đưa ra một phiên bản tối ưu hóa của điều này [2]. Gavrila & Philomen [8] thực hiện một cách tiếp cận trực tiếp hơn, trích xuất các hình ảnh cạnh và khớp chúng với một tập hợp các ví dụ đã học bằng cách sử dụng khoảng cách vát. Điều này đã được sử dụng trong một hệ thống phát hiện người đi bộ thời gian thực thực tế [7]. vi-ô-lông*et al*[22] xây dựng một bộ phát hiện người đang chuyển động hiệu quả, sử dụng AdaBoost để đào tạo một chuỗi các quy tắc loại bỏ vùng ngày càng phức tạp hơn dựa trên các bước sóng giống như Haar và sự khác biệt về không-thời gian. Ronfard*et al*[19] xây dựng một máy dò cơ thể có khớp nối bằng cách kết hợp các bộ phân loại chi dựa trên SVM trên1stvà2thứsắp xếp các bộ lọc Gaussian trong một khung lập trình động tương tự như của Felzenszwalb & Huttenlocher [3] và Ioffe & Forsyth [9]. Mikolajczyk*et al*[16] sử dụng kết hợp các biểu đồ định hướng định hướng với cường độ gradient ngưỡng nhị phân để xây dựng một phương pháp dựa trên các bộ phận có chứa các bộ phát hiện khuôn mặt, đầu và mặt trước và mặt bên của các bộ phận trên và dưới cơ thể. Ngược lại, máy dò của chúng tôi sử dụng kiến trúc đơn giản hơn với một cửa sổ phát hiện duy nhất, nhưng dường như mang lại hiệu suất cao hơn đáng kể đối với hình ảnh người đi bộ.

3 Tổng quan về phương pháp

Phần này cung cấp tổng quan về chuỗi trích xuất tính năng của chúng tôi, được tóm tắt trong hình. 1. Chi tiết thực hiện được hoãn lại cho đến khi§6. Phương pháp này dựa trên việc đánh giá các biểu đồ cục bộ được chuẩn hóa tốt của hướng gradient hình ảnh trong một lưới dày đặc. Các tính năng tương tự đã được sử dụng ngày càng nhiều trong thập kỷ qua [4, 5, 12, 15]. Ý tưởng cơ bản là diện mạo và hình dạng của đối tượng cục bộ thường có thể được mô tả khá tốt bằng sự phân bố của các gradient cường độ cục bộ hoặc



Hình 1. Tổng quan về chuỗi trích xuất tính năng và phát hiện đối tượng của chúng tôi. Cửa sổ máy dò được xếp bằng một lưới các khối chồng lên nhau trong đó Biểu đồ của các vectơ đặc trưng Dải màu định hướng được trích xuất. Các vectơ kết hợp được đưa vào một SVM tuyến tính để phân loại đối tượng/không phải đối tượng. Cửa sổ phát hiện được quét qua hình ảnh ở tất cả các vị trí và tỷ lệ, và triệt tiêu không tối đa thông thường được chạy trên kim tự tháp đầu ra để phát hiện các trường hợp đối tượng, nhưng bài báo này tập trung vào quy trình trích xuất tính năng.

hướng cạnh, ngay cả khi không có kiến thức chính xác về độ dốc hoặc vị trí cạnh tương ứng. Trong thực tế, điều này được thực hiện bằng cách chia cửa sổ hình ảnh thành các vùng không gian nhỏ ("*tế bào*"), đối với mỗi ô tích lũy biểu đồ 1-D cục bộ của hướng dốc hoặc hướng cạnh trên các pixel của ô. Các mục nhập biểu đồ kết hợp tạo thành đại diện. Để có sự bất biến tốt hơn đối với chiếu sáng, đổ bóng,*vân vân*., nó cũng hữu ích để chuẩn hóa tương phản các phản hồi cục bộ trước khi sử dụng chúng. Điều này có thể được thực hiện bằng cách tích lũy một phép đo "năng lượng" biểu đồ cục bộ trên các vùng không gian lớn hơn một chút ("*khối*") và sử dụng kết quả để chuẩn hóa tất cả các ô trong khối. Chúng tôi sẽ đề cập đến các khối mô tả chuẩn hóa như*Biểu đồ của Gradient định hướng (HOG)*bộ mô tả. Việc sắp xếp cửa sổ phát hiện bằng một lưới các bộ mô tả HOG dày đặc (trên thực tế, chồng chéo) và sử dụng vectơ đặc trưng kết hợp trong bộ phân loại cửa sổ dựa trên SVM thông thường mang lại chuỗi phát hiện con người của chúng tôi (xem hình 1).

Việc sử dụng các biểu đồ định hướng có nhiều tiền thân [13, 4, 5], nhưng nó chỉ đạt đến độ chín muồi khi được kết hợp với biểu đồ không gian cục bộ và chuẩn hóa trong Lowe's *Chuyển đổi tính năng bất biến tỷ lệ (SIFT)*cách tiếp cận đối sánh hình ảnh cơ sở rộng [12], trong đó nó cung cấp bộ mô tả bản vá hình ảnh cơ bản để khớp các điểm khóa bất biến tỷ lệ. Cách tiếp cận kiểu SIFT thực hiện rất tốt trong ứng dụng này [12, 14]. Các*Bối cảnh hình dạng* work [1] đã nghiên cứu các hình dạng khối và ô thay thế, mặc dù ban đầu chỉ sử dụng số lượng pixel cạnh mà không có biểu đồ định hướng làm cho biểu diễn trở nên hiệu quả. Sự thành công của các biểu diễn dựa trên tính năng thưa thớt này đã phần nào làm lu mờ sức mạnh và sự đơn giản của HOG với tư cách là các bộ mô tả hình ảnh dày đặc. Chúng tôi hy vọng rằng nghiên cứu của chúng tôi sẽ giúp khắc phục điều này. Đặc biệt, các thử nghiệm không chính thức của chúng tôi cho thấy rằng ngay cả các phương pháp tiếp cận dựa trên điểm chính tốt nhất hiện tại cũng có khả năng có tỷ lệ dương tính giả cao hơn ít nhất 1–2 bậc độ lớn so với phương pháp tiếp cận dạng lưới dày đặc của chúng tôi để phát hiện con người, chủ yếu là do không có công cụ phát hiện điểm chính nào mà chúng tôi đang sử dụng biết phát hiện cấu trúc cơ thể con người một cách đáng tin cậy.

Biểu diễn HOG/SIFT có một số ưu điểm. Nó nắm bắt cấu trúc cạnh hoặc độ dốc rất đặc trưng của hình dạng cục bộ và nó làm như vậy trong một biểu diễn cục bộ với mức độ bất biến có thể kiểm soát dễ dàng đối với các phép biến đổi hình học và trắc quang cục bộ: các phép tịnh tiến hoặc phép quay tạo ra ít khác biệt nếu chúng nhỏ hơn nhiều so với cấu trúc cục bộ. kích thước thùng định hướng hoặc không gian. Để phát hiện con người, thay vì lấy mẫu không gian thô, lấy mẫu định hướng tinh và chuẩn hóa trắc quang cục bộ hóa ra là chiến lược tốt nhất, có lẽ vì nó cho phép các chi và các đoạn cơ thể thay đổi hình dáng bên ngoài và di chuyển từ bên này sang bên kia khá nhiều với điều kiện là chúng duy trì hướng gần như thẳng đứng.

4 Tập dữ liệu và Phương pháp luận

Bộ dữ liệu. Chúng tôi đã thử nghiệm máy dò của mình trên hai bộ dữ liệu khác nhau. Đầu tiên là cơ sở dữ liệu về người đi bộ của MIT [18] đã được thiết lập tốt, chứa 509 hình ảnh đào tạo và 200 hình ảnh thử nghiệm về người đi bộ trong cảnh thành phố (cộng với phản xạ trái-phải của những hình ảnh này). Nó chỉ chứa các chế độ xem trước hoặc sau với một số tư thế tương đối hạn chế. Các máy dò tốt nhất của chúng tôi cho kết quả về cơ bản là hoàn hảo trên tập dữ liệu này, vì vậy chúng tôi đã tạo ra một tập dữ liệu mới và thách thức hơn đáng kể, 'INRIA', chứa 180564×128hình ảnh của con người được cắt từ một bộ ảnh cá nhân khác nhau. Hình 2 cho thấy một số mẫu. Mọi người thường đứng, nhưng xuất hiện theo bất kỳ hướng nào và dựa trên nhiều hình nền bao gồm cả đám đông. Nhiều người là người ngoài cuộc được chụp từ nền ảnh, vì vậy không có sự thiên vị cụ thể nào về tư thế của họ. Cơ sở dữ liệu có sẵn từ *http://lear.inrialpes.fr/data*cho mục đích nghiên cứu.

phương pháp luận. Chúng tôi đã chọn 1239 hình ảnh làm ví dụ đào tạo tích cực, cùng với phản xạ từ trái sang phải của chúng (tổng cộng 2478 hình ảnh). Một bộ cố định gồm 12180 bản vá được lấy mẫu ngẫu nhiên từ 1218 ảnh đào tạo không có người cung cấp bộ tiêu cực ban đầu. Đối với mỗi tổ hợp máy dò và tham số, một máy dò sơ bộ được đào tạo và 1218 ảnh đào tạo âm bản được tìm kiếm một cách thấu đáo để tìm các kết quả dương tính giả ('ví dụ khó'). Sau đó, phương pháp này được đào tạo lại bằng cách sử dụng bộ tăng cường này (12180 ban đầu + ví dụ khó) để tạo ra bộ phát hiện cuối cùng. Tập hợp các ví dụ cứng được lấy mẫu phụ nếu cần, sao cho các bộ mô tả của tập huấn luyện cuối cùng phù hợp với 1,7 Gb RAM để huấn luyện SVM. Quá trình đào tạo lại này cải thiện đáng kể hiệu suất của từng máy dò (5% tại10^{-4}Kết quả dương tính giả trên mỗi cửa sổ được kiểm tra (FPPW) cho trình phát hiện mặc định của chúng tôi), nhưng các vòng đào tạo lại bổ sung tạo ra một chút khác biệt nên chúng tôi không sử dụng chúng.

Để định lượng hiệu suất của máy dò, chúng tôi vẽ sơ đồ*Đánh đổi lỗi phát hiện (DET)*các đường cong trên thang đo log-log,*I E*. tỷ lệ bỏ lỡ (1−Nhớ lại hoặc~~TruePos/FalseNeg~~Sai phủ định) so với FPPW. Giá trị thấp hơn là tốt hơn. Biểu đồ DET được sử dụng rộng rãi trong lời nói và trong các đánh giá của NIST. Chúng trình bày thông tin tương tự như Đặc điểm hoạt động của máy thu (ROC's) nhưng cho phép nhỏ

Hình 2. Một số hình ảnh mẫu từ cơ sở dữ liệu phát hiện con người mới của chúng tôi. Đối tượng luôn ở tư thế thẳng đứng, nhưng có một số chỗ bị che khuất một phần và có nhiều thay đổi về tư thế, diện mạo, quần áo, ánh sáng và hậu cảnh.

xác suất được phân biệt dễ dàng hơn. Chúng tôi sẽ thường sử dụng tỷ lệ bỏ lỡ tại10^{-4}FPPW làm điểm tham chiếu cho kết quả. Điều này là tùy ý nhưng không nhiều hơn,*ví dụ*. Khu vực dưới ROC. Trong máy dò đa thang đo, nó tương ứng với tỷ lệ lỗi thô khoảng0,8dương tính giả trên mỗi640×480hình ảnh được kiểm tra. (Máy dò đầy đủ có tỷ lệ dương tính giả thậm chí còn thấp hơn do triệt tiêu không tối đa). Các đường cong DET của chúng tôi thường khá nông, do đó, ngay cả những cải thiện rất nhỏ về tỷ lệ bỏ lỡ cũng tương đương với mức tăng lớn trong FPPW với tỷ lệ bỏ lỡ không đổi. Ví dụ: đối với trình phát hiện mặc định của chúng tôi ở 1e-4 FPPW, cứ giảm 1% tuyệt đối (9% tương đối) tỷ lệ trượt tương đương với việc giảm FPPW ở tỷ lệ trượt không đổi theo hệ số 1,57.

5 Tổng quan về kết quả

Trước khi trình bày phân tích hiệu suất và triển khai chi tiết của chúng tôi, chúng tôi so sánh hiệu suất tổng thể của các máy dò HOG cuối cùng của chúng tôi với hiệu suất của một số phương pháp hiện có khác. Máy dò dựa trên khối hình chữ nhật (R-HOG) hoặc khối log-cực tròn (C-HOG) và SVM tuyến tính hoặc nhân được so sánh với các triển khai của chúng tôi về phương pháp tiếp cận bối cảnh hình dạng và sóng con Haar. Tóm lại, các cách tiếp cận này như sau:

Wavelet Haar tổng quát.Đây là một tập mở rộng của các wavelet giống Haar định hướng tương tự (nhưng tốt hơn) được sử dụng trong [17]. Các tính năng là phản ứng sửa chữa từ9×9và 12×12định hướng1stvà2^{thứ}bộ lọc hộp phái sinh tại45°khoảng và tương ứng2^{thứ}phát sinhxy filọc.

PCA-SIFT.Các bộ mô tả này dựa trên việc chiếu các hình ảnh chuyển màu trên cơ sở học được từ các hình ảnh đào tạo bằng PCA [11]. Ke & Sukthankar nhận thấy rằng họ vượt trội so với SIFT đối với kết hợp dựa trên điểm chính, nhưng điều này đang gây tranh cãi [14]. triển khai của chúng tôi sử dụng16×16các khối có cùng tỷ lệ đạo hàm, trùng nhau,*vân vân.*, cài đặt làm bộ mô tả HOG của chúng tôi. Cơ sở PCA được tính toán bằng cách sử dụng hình ảnh đào tạo tích cực.

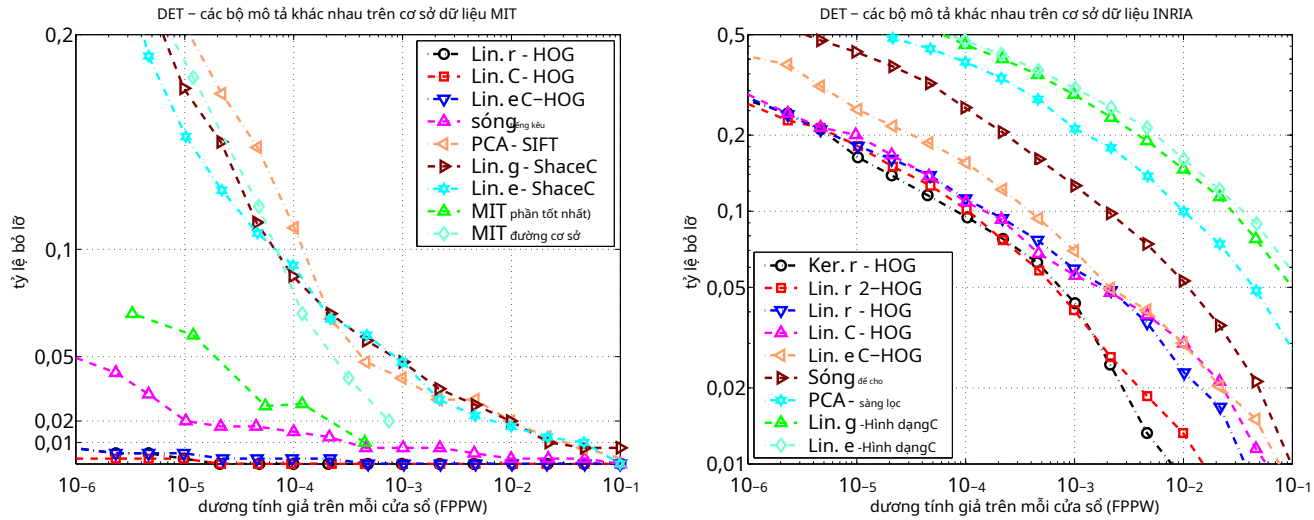
Bối cảnh hình dạng.Bối cảnh hình dạng ban đầu [1] đã sử dụng biểu quyết hiện diện cạnh nhị phân vào các thùng cách nhau log-cực, bất kể hướng cạnh. Chúng tôi mô phỏng điều này bằng bộ mô tả C-HOG (xem bên dưới) chỉ với 1 ngăn định hướng. 16 khoảng góc và 3 khoảng bán kính với bán kính trong 2 pixel và bán kính ngoài 8 pixel cho kết quả tốt nhất. Cả hai gradient-Biểu quyết dựa trên cường độ và sự hiện diện của cạnh đã được kiểm tra, với ngưỡng cạnh được chọn tự động để tối đa hóa hiệu suất phát hiện (các giá trị được chọn có phần thay đổi, trong vùng 20–50 cấp độ xám).

Kết quả.Hình 3 cho thấy hiệu suất của các máy dò khác nhau trên bộ dữ liệu MIT và INRIA. Các máy dò dựa trên HOG vượt trội hơn rất nhiều so với các máy dò wavelet, PCA-SIFT và Shape Context, mang lại sự phân tách gần như hoàn hảo trên bộ thử nghiệm MIT và ít nhất là một thứ tự giảm cường độ trong FPPW trên bộ INRIA. Các wavelet giống Haar của chúng tôi hoạt động tốt hơn các wavelet MIT bởi vì chúng tôi cũng sử dụng2^{thứ}đạo hàm thứ tự và độ tương phản chuẩn hóa vector đầu ra. Hình 3(a) cũng cho thấy các máy dò nguyên khối và dựa trên các bộ phận tốt nhất của MIT (các điểm được nội suy từ [17]), tuy nhiên hãy lưu ý rằng không thể so sánh chính xác vì chúng tôi không biết cách cơ sở dữ liệu trong [17] được chia thành các phần đào tạo và kiểm tra và các hình ảnh tiêu cực được sử dụng không có sẵn. Hiệu suất của máy dò hình chữ nhật (R-HOG) và hình tròn (C-HOG) cuối cùng rất giống nhau, với C-HOG có lợi thế hơn một chút. Tăng cường R-HOG với các máy dò thanh nguyên thủy (được định hướng2^{thứ} dẫn xuất – 'R2-HOG') tăng gấp đôi kích thước tính năng nhưng cải thiện hơn nữa hiệu suất (2% tại10^{-4}FPPW). Thay thế SVM tuyến tính bằng hạt nhân Gaussian sẽ cải thiện hiệu suất khoảng 3% tại10^{-4}FPPW, với chi phí thời gian chạy cao hơn nhiều[1]. Sử dụng biểu quyết cạnh nhị phân (EC-HOG) thay vì biểu quyết có trọng số độ dốc (C-HOG) làm giảm hiệu suất 5% tại10^{-4}FPPW, trong khi bỏ qua thông tin định hướng, nó sẽ giảm nhiều hơn nữa, ngay cả khi thêm các ngăn không gian hoặc hướng tâm bổ sung (33% tại 10^{-4}FPPW, cho cả hai cạnh (E-ShapeC) và độ dốc (G-ShapeC)). PCA-SIFT cũng hoạt động kém. Một lý do là, so với [11], nhiều vectơ chính hơn (80 trên 512) phải được giữ lại để thu được cùng một tỷ lệ phương sai. Điều này có thể là do đăng ký không gian yếu hơn khi không có bộ phát hiện điểm chính.

6 Nghiên cứu Thực hiện và Thực hiện

Bây giờ chúng tôi cung cấp chi tiết về việc triển khai HOG của chúng tôi và nghiên cứu một cách có hệ thống tác động của các lựa chọn khác nhau đối với

[1]Chúng tôi sử dụng các ví dụ khó được tạo bởi*tuyến tính*R-HOG để huấn luyện trình phát hiện R-HOG nhân, vì R-HOG nhân tạo ra quá ít kết quả dương tính giả nên bộ ví dụ cứng của nó quá thưa thớt để cải thiện đáng kể khả năng tổng quát hóa.



Hình 3. Hiệu suất của các máy dò được chọn trên bộ dữ liệu MIT (trái) và (phải) INRIA. Xem các văn bản để biết chi tiết.

hiệu suất tector. Trong suốt phần này, chúng tôi đề cập đến kết quả cho trình phát hiện mặc định của chúng tôi có các thuộc tính sau, được mô tả bên dưới: Không gian màu RGB không có hiệu chỉnh gamma; [-1, 0, 1]bộ lọc gradient không làm mịn; biểu quyết độ dốc tuyến tính vào 9 ngăn định hướng trong0◦–180◦;16×16 khối pixel của bốnsố 8×số 8ô điểm ảnh; Cửa sổ không gian Gauss với σ =số 8điểm ảnh;*L2-Hys*(Lowe-style clipped L2 norm) chuẩn hóa khối; khoảng cách giữa các khối là 8 pixel (do đó độ bao phủ của mỗi ô gấp 4 lần); 64×128cửa sổ phát hiện; bộ phân loại SVM tuyến tính.

Hình 4 tóm tắt tác động của các tham số HOG khác nhau đối với hiệu suất phát hiện tổng thể. Những điều này sẽ được xem xét chi tiết dưới đây. Các kết luận chính là để có hiệu suất tốt, người ta nên sử dụng các công cụ phái sinh tỷ lệ nhỏ (về cơ bản là không làm mịn), nhiều ngăn định hướng và các khối mô tả chồng chéo, chuẩn hóa mạnh, có kích thước vừa phải.

6.1 Bình thường hóa Gamma/Màu sắc

Chúng tôi đã đánh giá một số biểu diễn pixel đầu vào bao gồm các không gian màu thang độ xám, RGB và LAB tùy chọn với cân bằng luật lũy thừa (gamma). Những chuẩn hóa này chỉ có tác động khiêm tốn đến hiệu suất, có lẽ vì chuẩn hóa bộ mô tả tiếp theo đạt được kết quả tương tự. Chúng tôi sử dụng thông tin màu sắc khi có sẵn. Không gian màu RGB và LAB cho kết quả có thể so sánh được, nhưng việc hạn chế đối với thang độ xám sẽ làm giảm hiệu suất 1,5% tại10^{-4} FPPW. Nén gamma căn bậc hai của mỗi kênh màu cải thiện hiệu suất ở FPPW thấp (1% ở10^{-4}FPPW) nhưng nén nhật ký quá mạnh và làm trầm trọng thêm 2% tại 10^{-4}FPPW.

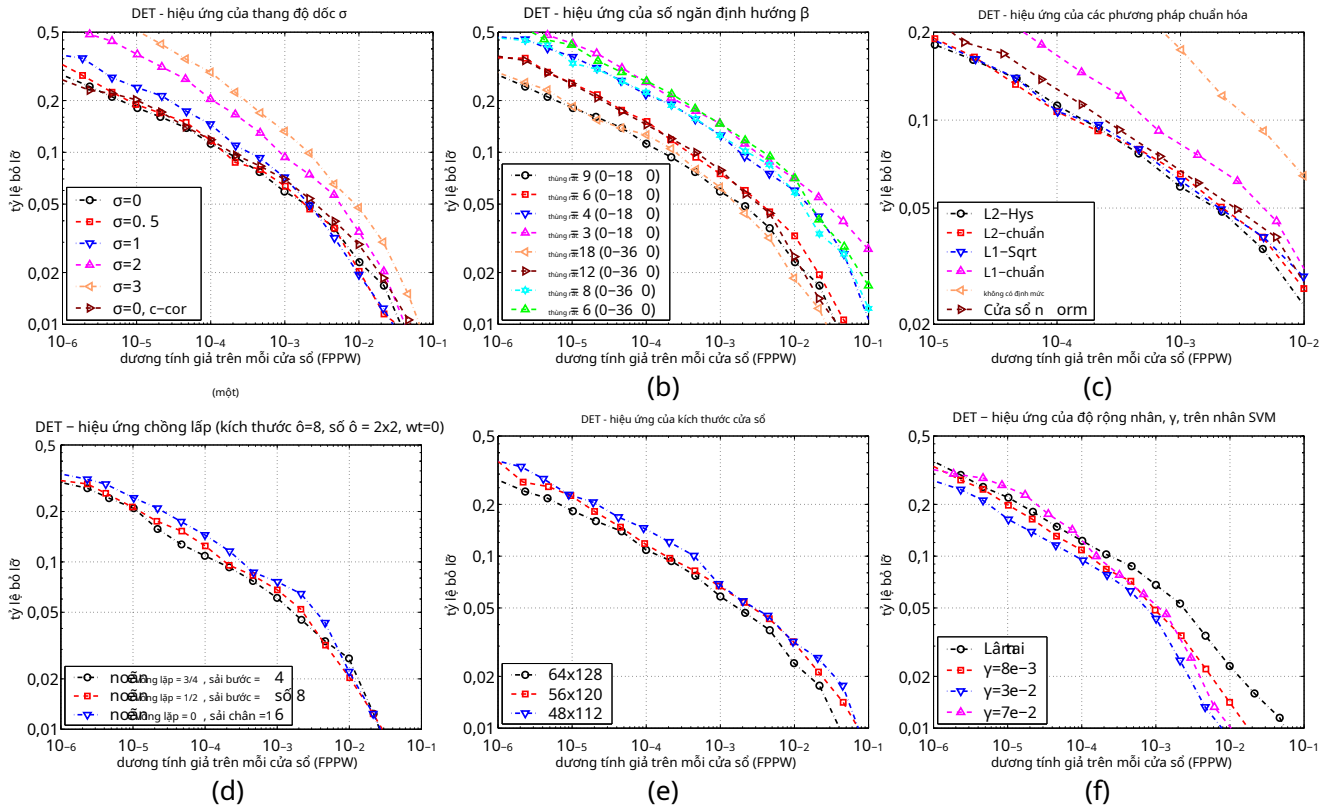
6.2 Tính toán độ dốc

Hiệu suất của máy dò nhạy cảm với cách tính toán độ dốc, nhưng sơ đồ đơn giản nhất hóa ra lại là tốt nhất. Chúng tôi đã thử nghiệm các gradient được tính toán bằng cách sử dụng làm mịn Gaussian, sau đó là một trong số các dẫn xuất rời rạc. mặt nạ tive. Một số thang đo làm mịn đã được thử nghiệm bao gồm σ=0 (không ai). Các mặt nạ được thử nghiệm bao gồm các dẫn xuất điểm 1-D khác nhau (không tập trung [-1, 1],căn giữa [−1, 0, 1]và lập phương [1, −8, 0, (8, −1])cũng như3×3Mặt nạ Sobel và 2×2đường chéo$\begin{smallmatrix}-1 & 0\\ 0 & 1\end{smallmatrix}$, $\begin{smallmatrix}0 & 1\\ -1 & 0\end{smallmatrix}$ (xây dựng nhỏ gọn nhất tred mặt nạ phái sinh 2-D). 1-D đơn giản [−1, 0, 1]mặt nạ tại σ=0làm việc một cách tốt nhất. Sử dụng mặt nạ lớn hơn dường như luôn làm giảm hiệu suất và việc làm mịn sẽ làm hỏng nó một cách đáng kể: đối với các dẫn xuất Gaussian, di chuyển từ σ=0đến σ=2giảm tỷ lệ thu hồi từ 89% xuống 80% tại10^{-4}FPPW. Tại σ=0, các bộ lọc 5 chiều rộng 1-D được hiệu chỉnh khối kém hơn khoảng 1% so với [-1, 0, 1]tại10^{-4}FPPW, trong khi2×2mặt nạ chéo kém hơn 1,5%. Sử dụng [-1, 1]mặt nạ phái sinh cũng làm giảm hiệu suất (1,5% tại10^{-4}FPPW), có lẽ là do ước tính định hướng bị ảnh hưởng do xvày filters được đặt tại các trung tâm khác nhau.

Đối với hình ảnh màu, chúng tôi tính toán độ dốc riêng biệt cho từng kênh màu và lấy kênh có chuẩn lớn nhất làm vectơ độ dốc của pixel.

6.3 Ghép không gian/định hướng

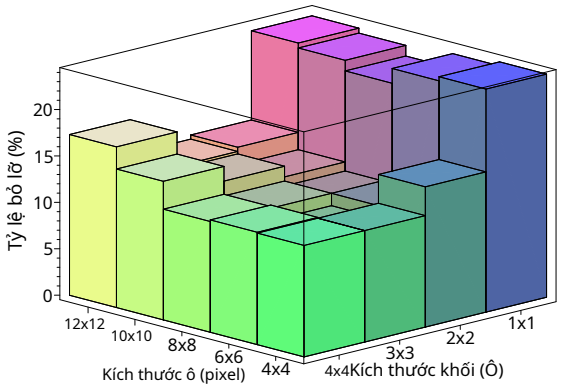
Bước tiếp theo là tính phi tuyến cơ bản của bộ mô tả. Mỗi pixel tính toán một phiếu bầu có trọng số cho một kênh biểu đồ hướng cạnh dựa trên hướng của phần tử chuyển màu tập trung vào nó và các phiếu bầu được tích lũy vào các ngăn định hướng trên các vùng không gian cục bộ mà chúng tôi gọi là*tế bào*. Các ô có thể là hình chữ nhật hoặc hình tròn (các cung cực log). Các ngăn định hướng được đặt cách đều nhau trên0◦– 180◦(“độ dốc không dấu”) hoặc0◦–360◦(“đã ký” độ dốc). Để giảm răng cưa, các phiếu bầu được nội suy song tuyến giữa các trung tâm thùng lân cận theo cả hướng và vị trí. Biểu quyết là một hàm của cường độ gradient tại pixel, hoặc là độ lớn của chính nó, bình phương của nó, căn bậc hai của nó hoặc dạng cắt bớt của cường độ biểu thị sự hiện diện/không có cạnh mềm của một pixel tại pixel. Trong thực tế, sử dụng các



Hình 4. Để biết chi tiết, xem văn bản. (a) Sử dụng thang đo đạo hàm tốt làm tăng đáng kể hiệu suất. ('c-cor' là đạo hàm điểm hiệu chỉnh bậc ba 1D). (b) Việc tăng số lượng ngăn định hướng giúp tăng hiệu suất đáng kể lên tới khoảng 9 ngăn được đặt cách nhau trên0◦– 180◦. (c) Ảnh hưởng của các lược đồ chuẩn hóa khối khác nhau (xem§6.4). (d) Sử dụng các khối mô tả chồng chéo làm giảm tỷ lệ bỏ sót khoảng 5%. (e) Giảm lề 16 pixel xung quanh64×128cửa sổ phát hiện làm giảm hiệu suất khoảng 3%. (f) Sử dụng SVM hạt nhân Gaussian,$\exp(-\gamma\|x_1-x_2\|_2)$, cải thiện hiệu suất khoảng 3%.

cường độ chính nó cho kết quả tốt nhất. Lấy căn bậc hai làm giảm hiệu suất một chút, trong khi sử dụng biểu quyết hiện diện cạnh nhị phân làm giảm hiệu suất đáng kể (5% tại10^{-4}FPPW).

Mã hóa định hướng tốt hóa ra là điều cần thiết để có hiệu suất tốt, trong khi đó (xem bên dưới) việc tạo thùng theo không gian có thể khá thô. Như hình. 4(b) cho thấy, việc tăng số lượng ngăn định hướng giúp cải thiện hiệu suất đáng kể lên đến khoảng 9 ngăn, nhưng không tạo ra nhiều khác biệt ngoài điều này. Cái này dành cho các thùng cách nhau trên0◦–180◦,*I.E.* 'ký hiệu' của gradient bị bỏ qua. Bao gồm các gradient đã ký (phạm vi định hướng 0◦–360◦, như trong bộ mô tả SIFT ban đầu) làm giảm hiệu suất, ngay cả khi số lượng ngăn cũng tăng gấp đôi để duy trì độ phân giải hướng ban đầu. Đối với con người, nhiều loại quần áo và màu nền có lẽ làm cho các dấu hiệu tương phản trở nên không có thông tin. Tuy nhiên, lưu ý rằng việc bao gồm thông tin ký hiệu sẽ giúp ích đáng kể trong một số nhiệm vụ nhận dạng đối tượng khác,*ví dụ.* ô tô, xe máy.



Hình 5. Tỷ lệ bỏ lỡ tại10^{-4}FPPW khi kích thước ô và khối thay đổi. Bước tiến (khối chồng lên nhau) được cố định ở một nửa kích thước khối. 3×3 khối của6×6các ô pixel hoạt động tốt nhất với tỷ lệ trượt 10,4%.

6.4 Khối chuẩn hóa và mô tả

Cường độ chuyển màu thay đổi trong một phạm vi rộng do các biến thể cục bộ về độ sáng và độ tương phản nền trước-nền, do đó, việc chuẩn hóa độ tương phản cục bộ hiệu quả hóa ra lại rất cần thiết để có hiệu suất tốt. Chúng tôi đã đánh giá một số ber của các chương trình chuẩn hóa khác nhau. Hầu hết chúng dựa trên việc nhóm các ô thành các khối không gian lớn hơn và tương phản chuẩn hóa từng khối riêng biệt. Sau đó, bộ mô tả cuối cùng là vectơ của tất cả các thành phần của phản hồi ô đã chuẩn hóa từ tất cả các khối trong cửa sổ phát hiện.

Trên thực tế, chúng tôi thường chồng lấp các khối để mỗi phản hồi ô vô hướng đóng góp một số thành phần vào vectơ mô tả cuối cùng, mỗi thành phần được chuẩn hóa đối với một khối khác nhau. Điều này có vẻ dư thừa nhưng việc chuẩn hóa tốt là rất quan trọng và bao gồm cả chồng chéo sẽ cải thiện đáng kể hiệu suất. Hình 4(d) cho thấy hiệu suất tăng 4% khi10^{-4}FPPW khi chúng tôi tăng chồng lấp từ không có (sải 16) lên 16 lần diện tích / 4 lần phạm vi tuyến tính (sải 4).

Chúng tôi đã đánh giá hai lớp hình học khối, hình vuông hoặc hình chữ nhật được phân chia thành các lưới ô không gian hình vuông hoặc hình chữ nhật và các khối hình tròn được phân chia thành các ô theo kiểu log-cực. Chúng tôi sẽ gọi hai cách sắp xếp này là R-HOG và C-HOG (đối với HOG hình chữ nhật và hình tròn).

R-HOG.Các khối R-HOG có nhiều điểm tương đồng với các bộ mô tả SIFT [12] nhưng chúng được sử dụng khá khác nhau. Chúng được tính toán trong các lưới dày đặc ở một tỷ lệ duy nhất mà không căn chỉnh hướng chi phối và được sử dụng như một phần của vectơ mã lớn hơn mã hóa hoàn toàn vị trí không gian so với cửa sổ phát hiện, trong khi SIFT được tính toán ở một tập hợp thưa thớt các điểm chính bất biến theo tỷ lệ, được xoay để sắp xếp các hướng chi phối của chúng và được sử dụng riêng lẻ. SIFT được tối ưu hóa để khớp đường cơ sở thưa thớt, R-HOG dành cho mã hóa dạng không gian mạnh mẽ dày đặc. Các tiền thân khác bao gồm biểu đồ hướng cạnh của Freeman & Roth [4]. Chúng tôi thường sử dụng R-HOG vuông,*I E* . ς×ς lưới của η×η mỗi ô pixel chứa β ngăn định hướng, trong đó ς, η, β là các tham số.

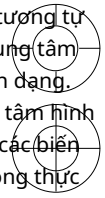
Hình. 5 biểu đồ tỷ lệ bỏ lỡ tại10^{-4}FPPW wrt kích thước ô và kích thước khối trong các ô. Để phát hiện con người,3×3khối tế bào của6×6các ô pixel hoạt động tốt nhất, với tỷ lệ trượt 10,4% ở 10^{-4}FPPW. Trên thực tế, các ô rộng 6–8 pixel hoạt động tốt nhất bất kể kích thước khối – một sự trùng hợp thú vị vì các chi của con người có chiều ngang khoảng 6–8 pixel trong hình ảnh của chúng tôi.2×2và 3×3khối hoạt động tốt nhất. Ngoài điều này, kết quả xấu đi: khả năng thích ứng với các điều kiện hình ảnh cục bộ bị suy yếu khi khối trở nên quá lớn và khi nó quá nhỏ (1×1 chặn/chuẩn hóa theo hướng một mình) thông tin không gian có giá trị bị chặn.

Như trong [12], sẽ rất hữu ích khi giảm trọng lượng các pixel gần các cạnh của khối bằng cách áp dụng cửa sổ không gian Gaussian cho từng pixel trước khi tích lũy các phiếu định hướng vào các ô. Điều này cải thiện hiệu suất thêm 1% tại10^{-4}FPPW cho một Gaussian với σ =0,5∗chiều rộng khối.

Chúng tôi cũng đã thử bao gồm nhiều loại khối với các kích thước ô và khối khác nhau trong bộ mô tả tổng thể. Điều này cải thiện một chút hiệu suất (khoảng 3% tại10^{-4}FPPW), với cái giá phải trả là kích thước bộ mô tả tăng lên rất nhiều.

Bên cạnh các khối R-HOG vuông, chúng tôi cũng đã thử nghiệm các khối dọc (2×1ô) và ngang (1×2ô) và một bộ mô tả kết hợp bao gồm cả cặp dọc và ngang. Các cặp dọc và dọc+ngang tốt hơn đáng kể so với chỉ các cặp ngang, nhưng không tốt bằng2×2khối (kém hơn 1% ở10^{-4}FPPW).

C-HOG.Các bộ mô tả khối tròn (C-HOG) của chúng tôi gợi nhớ đến Ngữ cảnh hình dạng [1] ngoại trừ điều quan trọng là mỗi ô không gian chứa một chồng các ô định hướng có trọng số độ dốc thay vì một số lượng hiện diện cạnh độc lập với một hướng. Lưới phân cực logarit ban đầu được đề xuất bởi ý tưởng rằng nó sẽ cho phép kết hợp mã hóa tốt cấu trúc lân cận với mã hóa thô hơn của bối cảnh rộng hơn và thực tế là sự chuyển đổi từ trường thị giác sang vỏ não V1 ở loài linh trưởng là logarit [21]. Tuy nhiên, các bộ mô tả nhỏ với rất ít thùng hướng tâm hóa ra lại mang lại hiệu suất tốt nhất, vì vậy trong thực tế có rất ít sự không đồng nhất hoặc ngữ cảnh. Có lẽ tốt hơn là nghĩ về C-HOG đơn giản như một hình thức mã hóa bao quanh trung tâm tiên tiến.



Chúng tôi đã đánh giá hai biến thể của hình dạng C-HOG, một biến thể có một ô trung tâm hình tròn duy nhất (tương tự như tính năng GLOH của [14]) và một biến thể có ô trung tâm được chia thành các cung góc như trong bối cảnh hình dạng. Chúng tôi chỉ trình bày kết quả cho các biến thể trung tâm hình tròn, vì các biến thể này có ít ô không gian hơn so với các biến thể trung tâm được chia và cho hiệu suất tương tự trong thực tế. Một báo cáo kỹ thuật sẽ cung cấp thêm chi tiết. Bố cục C-HOG có bốn tham số: số lượng ngăn góc và hướng tâm; bán kính của thùng trung tâm tính bằng pixel; và hệ số mở rộng cho các bán kính tiếp theo. Cần có ít nhất hai ngăn hướng tâm (trung tâm và xung quanh) và bốn ngăn góc (chia tư) để có hiệu suất tốt. Bao gồm các ngăn hướng tâm bổ sung không làm thay đổi nhiều hiệu suất, trong khi việc tăng số lượng ngăn góc sẽ làm giảm hiệu suất (1,3% tại10^{-4}FPPW khi đi từ 4 đến 12 thùng góc). 4 pixel là bán kính tốt nhất cho thùng trung tâm, nhưng 3 và 5 cho kết quả tương tự. Việc tăng hệ số mở rộng từ 2 lên 3 khiến hiệu suất về cơ bản không thay đổi. Với các tham số này, cả trọng số không gian Gaussian và trọng số nghịch đảo của phiếu bầu ô theo vùng ô đều không làm thay đổi hiệu suất, nhưng việc kết hợp hai điều này sẽ giảm nhẹ. Các giá trị này giả sử lấy mẫu định hướng tốt. Bối cảnh hình dạng (1 thùng định hướng) yêu cầu phân chia không gian tốt hơn nhiều để hoạt động tốt.

Đề án chuẩn hóa khối. Chúng tôi đã đánh giá bốn sơ đồ chuẩn hóa khối khác nhau cho từng dạng hình học HOG ở trên. Để chovlà vectơ mô tả không chuẩn hóa, $\|v\|_k$là của nók-định mức chok=1, 2,và ε là√ một hằng số nhỏ. Các kế hoạch là: (a)*L2-chuẩn*,v→v/$\|v\|_2$ $_2+\varepsilon_2$; (b) *L2-Hys*, L2-norm theo sau bởi clipping (giới hạn giá trị tối đa củavđến 0,2) và chuẩn hóa lại, như trong [12]; (c) *L1-chuẩn*,v→v/($\|v\|_1$+√ ε);và (d)*L1-sqrt*, L1-norm fol-giảm bởi căn bậc haiv→ v/($\|v\|_1$+ ε),số tiền nào để coi các vectơ mô tả là phân phối xác suất và sử dụng khoảng cách Bhattacharya giữa chúng. Hình 4(c) cho thấy rằng L2-Hys, L2-norm và L1-sqrt đều hoạt động tốt như nhau, trong khi L1-norm đơn giản làm giảm 5% hiệu suất và bỏ qua hoàn toàn chuẩn hóa sẽ giảm 27% hiệu suất, tại 10^{-4}FPPW. Một số chính quy hóa ε là cần thiết khi chúng tôi đánh giá-

đã sử dụng dày đặc các bộ mô tả, kể cả trên các mảng trống, nhưng kết quả không nhạy cảm với giá trị của ε trong một phạm vi lớn.

Chuẩn hóa xung quanh trung tâm. Chúng tôi cũng đã nghiên cứu một sơ đồ chuẩn hóa ô kiểu bao quanh trung tâm thay thế, trong đó hình ảnh được xếp bằng một lưới các ô và đối với mỗi ô, tổng năng lượng trong ô và vùng xung quanh của nó (được tổng hợp theo các hướng và gộp lại bằng cách sử dụng trọng số Gaussian) được sử dụng để bình thường hóa tế bào. Tuy nhiên như hình. 4(c) ("*định mức cửa sổ*") cho thấy, điều này làm giảm hiệu suất so với sơ đồ dựa trên khối tương ứng (2% tại10^{-4} FPPW, để gộp với σ=1chiều rộng ô). Một lý do là không còn bất kỳ khối chồng chéo nào nên mỗi ô chỉ được mã hóa một lần trong bộ mô tả cuối cùng. Bao gồm một số chuẩn hóa cho mỗi ô dựa trên các tỷ lệ tổng hợp khác nhau σkhông cung cấp thay đổi rõ rệt về hiệu suất, vì vậy có vẻ như đó là sự tồn tại của một số vùng tổng hợp với*khác nhau*độ lệch không gian so với ô quan trọng ở đây, không phải tỷ lệ tổng hợp.

Để làm rõ điểm này, hãy xem xét máy dò R-HOG với các khối chồng lên nhau. Các hệ số của SVM tuyến tính được đào tạo đưa ra thước đo xem mỗi ô của mỗi khối có thể có bao nhiêu trọng số trong quyết định phân biệt cuối cùng. Đóng kiểm tra hình. Hình 6(b,f) cho thấy các tế bào quan trọng nhất là những tế bào thường chứa các đường viền chính của con người (đặc biệt là đầu, vai và bàn chân), các khối viết chuẩn hóa nằm*ngoài*đường viền. Nói cách khác — mặc dù nền phức tạp, lộn xộn thường thấy trong tập huấn luyện của chúng tôi — tín hiệu của máy dò chủ yếu dựa trên độ tương phản của các đường viền bóng với nền, chứ không phải trên các cạnh bên trong hoặc trên các đường viền bóng so với nền trước. Quần áo có hoa văn và các biến thể tư thế có thể làm cho các vùng bên trong không đáng tin cậy như tín hiệu hoặc quá trình chuyển đổi từ tiền cảnh sang đường viền có thể bị nhầm lẫn bởi các hiệu ứng đổ bóng và đổ bóng mượt mà. Tương tự, hình. Hình 6(c,g) minh họa rằng độ dốc bên trong người (đặc biệt là độ dốc dọc) thường được tính là tín hiệu âm, có lẽ vì điều này triệt tiêu dương tính giả trong đó các đường thẳng đứng dài kích hoạt các tế bào đầu và chân theo chiều dọc.

6.5 Cửa sổ dò tìm và ngữ cảnh

Của chúng tôi64×128cửa sổ phát hiện bao gồm khoảng 16 pixel lề xung quanh người ở cả bốn phía. Hình 4(e) cho thấy đường viền này cung cấp một lượng ngữ cảnh đáng kể giúp phát hiện. Giảm nó từ 16 xuống 8 pixel (48×112cửa sổ phát hiện) giảm hiệu suất 6% tại10^{-4} FPPW. Giữ một64×128cửa sổ nhưng việc tăng kích thước người bên trong nó (lại giảm đường viền) gây ra sự giảm hiệu suất tương tự, mặc dù độ phân giải của người thực sự được tăng lên.

6.6 Bộ phân loại

Theo mặc định, chúng tôi sử dụng một mềm (c=0,01)SVM tuyến tính được đào tạo với SVMLight [10] (được sửa đổi một chút để giảm mức sử dụng bộ nhớ cho các sự cố với vectơ mô tả dày đặc lớn). Chúng ta-ing một nhân Gaussian SVM tăng hiệu suất khoảng 3% tại10^{-4}FPPW với chi phí thời gian chạy cao hơn nhiều.

6.7 Thảo luận

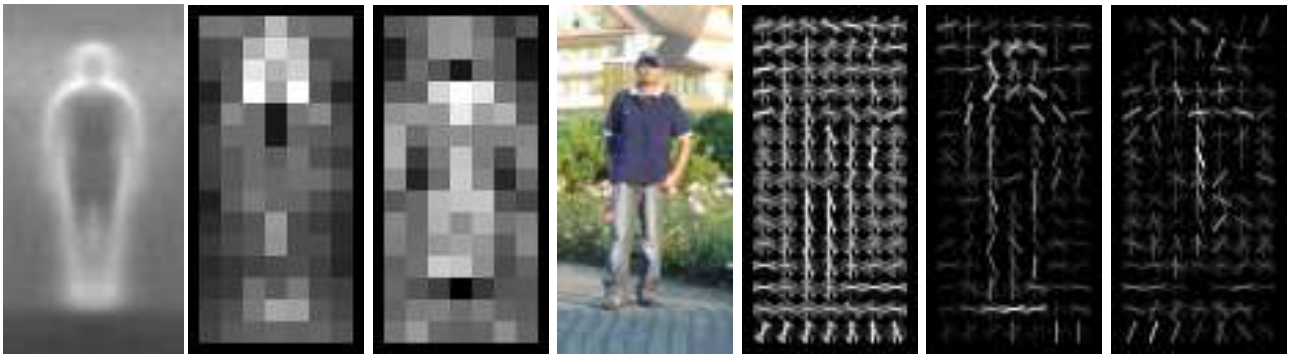
Nhìn chung, có một số phát hiện đáng chú ý trong công việc này. Thực tế là HOG hoạt động vượt trội hơn rất nhiều so với wavelet và bất kỳ mức độ làm mịn đáng kể nào trước khi tính toán độ dốc đều làm hỏng kết quả HOG nhấn mạnh rằng phần lớn thông tin hình ảnh có sẵn là từ*cạnh đột ngột ở quy mô tốt*và việc làm mờ điều này với hy vọng giảm độ nhạy đối với vị trí không gian là một sai lầm. Thay vào đó, độ dốc phải được tính toán ở tỷ lệ tốt nhất hiện có trong lớp kim tự tháp hiện tại, được chỉnh sửa hoặc sử dụng để biểu quyết định hướng và chỉ sau đó làm mờ không gian. Với điều này, đủ lượng tử hóa không gian tương đối thô (số 8×số 8ô pixel/chiều rộng một nhánh). Mặt khác, ít nhất là đối với khả năng phát hiện của con người, định hướng mẫu khá tinh vi: cả bối cảnh sóng con và hình dạng đều bị mất đi đáng kể ở đây.

Thứ hai, mạnh mẽ*địa phương*chuẩn hóa tương phản là điều cần thiết để có kết quả tốt và các sơ đồ kiểu trung tâm bao quanh truyền thống không phải là lựa chọn tốt nhất. Có thể đạt được kết quả tốt hơn bằng cách chuẩn hóa từng phần tử (cạnh, ô)*vài lần*đối với các hỗ trợ cục bộ khác nhau và coi kết quả là tín hiệu độc lập. Trong trình phát hiện tiêu chuẩn của chúng tôi, mỗi ô HOG xuất hiện bốn lần với các chuẩn hóa khác nhau và bao gồm cả thông tin 'dư thừa' này giúp cải thiện hiệu suất từ 84% lên 89% tại10^{-4}FPPW.

7 Tóm tắt và Kết luận

Chúng tôi đã chỉ ra rằng việc sử dụng biểu đồ định hướng độ dốc được chuẩn hóa cục bộ có các tính năng tương tự như bộ mô tả SIFT [12] trong một lưới chồng chéo dày đặc sẽ mang lại kết quả rất tốt cho việc phát hiện người, giảm tỷ lệ dương tính giả nhiều hơn một bậc độ lớn so với cơ sở sóng con Haar tốt nhất máy dò từ [17]. Chúng tôi đã nghiên cứu ảnh hưởng của các tham số mô tả khác nhau và kết luận rằng độ dốc tỷ lệ nhỏ, cách tạo ô định hướng tinh, cách tạo ô theo không gian tương đối thô và chuẩn hóa độ tương phản cục bộ chất lượng cao trong các khối mô tả chồng chéo đều quan trọng để có hiệu suất tốt. Chúng tôi cũng đã giới thiệu một cơ sở dữ liệu dành cho người đi bộ mới và thách thức hơn, được cung cấp công khai.

Công việc tương lai:Mặc dù trình phát hiện SVM tuyến tính hiện tại của chúng tôi khá hiệu quả – xử lý một320×240hình ảnh không gian tỷ lệ (4000 cửa sổ phát hiện) trong chưa đầy một giây – vẫn còn chỗ để tối ưu hóa và để tăng tốc độ phát hiện hơn nữa, sẽ rất hữu ích nếu phát triển một trình phát hiện kiểu chuỗi từ chối hoặc thô dựa trên các bộ mô tả HOG. Chúng tôi cũng đang nghiên cứu các máy dò dựa trên HOG kết hợp thông tin chuyển động bằng cách sử dụng khớp khối hoặc trường dòng quang. Cuối cùng, mặc dù máy dò kiểu mẫu cố định hiện tại tỏ ra khó bị đánh bại đối với những người đi bộ hoàn toàn có thể nhìn thấy, nhưng con người có khả năng khớp nối cao và chúng tôi tin rằng việc bao gồm một mô hình dựa trên các bộ phận với mức độ bất biến không gian cục bộ cao hơn



(một) (b) (c) (d) (e) (f) (g)

Hình 6. Máy dò HOG của chúng tôi chủ yếu dựa trên các đường viền của bóng (đặc biệt là đầu, vai và bàn chân). Các khối tích cực nhất là chỉ tập trung vào nền hình ảnh*ngoài*đường viền. (a) Hình ảnh độ dốc trung bình trên các ví dụ đào tạo. (b) Mỗi "pixel" hiển thị trọng số SVM dương tối đa trong khối có tâm là pixel. (c) Tương tự như vậy đối với các trọng số SVM âm. (d) Một hình ảnh thử nghiệm. (e) Đó là bộ mô tả R-HOG được tính toán. (f, g) Bộ mô tả R-HOG có trọng số tương ứng với trọng số SVM dương và âm.

sẽ giúp cải thiện kết quả phát hiện trong các tình huống tổng quát hơn.

Sự nhìn nhận.Công việc này được hỗ trợ bởi các dự án nghiên cứu của Liên minh Châu ÂuÁT CHỦMEDIAvà PASCAL. Chúng tôi cảm ơn Cordelia Schmid vì nhiều nhận xét hữu ích. SVM-Light [10] cung cấp khả năng huấn luyện SVM quy mô lớn đáng tin cậy.

Người giới thiệu

[1] S. Belongie, J. Malik và J. Puzicha. Hình dạng phù hợp.*ICCV lần thứ 8, Vancouver, Canada*, trang 454–461, 2001.

[2] V. de Poortere, J. Cant, B. Van den Bosch, J. de Prins, F. Fransens và L. Van Gool. Phát hiện người đi bộ hiệu quả: một trường hợp thử nghiệm để phân loại dựa trên svm. *Hội thảo về tầm nhìn nhận thức*, 2002. Có sẵn trực tuyến: http://www.vision.ethz.ch/cogvis02/.

[3] P. Felzenszwalb và D. Huttenlocher. Kết hợp hiệu quả các cấu trúc hình ảnh.*CVPR, Đảo Hilton Head, Nam Carolina, Hoa Kỳ*, trang 66–75, 2000.

[4] WT Freeman và M. Roth. Biểu đồ định hướng để nhận dạng cử chỉ tay.*quốc tế Hội thảo về Nhận dạng khuôn mặt và cử chỉ tự động, IEEE Computer Society, Zurich, Thụy Sĩ*, trang 296–301, tháng 6/1995.

[5] WT Freeman, K. Tanaka, J. Ohta và K. Kyuma. Thị giác máy tính cho trò chơi máy tính.*Hội nghị quốc tế lần thứ 2 về nhận dạng khuôn mặt và cử chỉ tự động, Killington, VT, Hoa Kỳ*, trang 100–105, tháng 10/1996.

[6] DM Gavrila. Phân tích trực quan về chuyển động của con người: Một cuộc khảo sát.*CVIU*, 73(1):82–98, 1999.

[7] DM Gavrila, J. Giebel, và S. Munder. Phát hiện người đi bộ dựa trên tầm nhìn: hệ thống bảo vệ+.*Proc. của Hội nghị chuyên đề về phương tiện thông minh IEEE, Parma, Ý*, 2004.

[8] Đ.M Gavrila và V. Philomin. Phát hiện đối tượng thời gian thực cho xe thông minh.*CVPR, Fort Collins, Colorado, Hoa Kỳ*, trang 87–93, 1999.

[9] S. Ioffe và DA Forsyth. Phương pháp xác suất để tìm người.*IJCV*, 43(1):45–68, 2001.

[10] T. Joachims. Làm cho việc học svm quy mô lớn trở nên thiết thực. Trong B. Schlkopf, C. Burges, và A. Smola, biên tập viên,*Những tiến bộ trong Kernel Methods - Hỗ trợ học Vector*. Nhà xuất bản MIT, Cambridge, MA, USA, 1999.

[11] Y. Ke và R. Sukthankar. Pca-sift: Một đại diện đặc biệt hơn cho các bộ mô tả hình ảnh cục bộ.*CVPR, Washington, DC, Hoa Kỳ*, trang 66–75, 2004.

[12] DG Lowe. Các tính năng hình ảnh đặc biệt từ các điểm chính không thay đổi tỷ lệ.*IJCV*, 60(2):91–110, 2004.

[13] RK McConnell. Phương pháp và thiết bị nhận dạng mẫu, tháng 1 năm 1986. Bằng sáng chế Hoa Kỳ số 4.567.610.

[14] K. Mikolajczyk và C. Schmid. Một đánh giá hiệu suất của các bộ mô tả địa phương.*PAMI*, 2004. Đã được chấp nhận.

[15] K. Mikolajczyk và C. Schmid. Bộ phát hiện điểm quan tâm bất biến theo tỷ lệ và affine.*IJCV*, 60(1):63–86, 2004.

[16] K. Mikolajczyk, C. Schmid, và A. Zisserman. Phát hiện con người dựa trên sự lắp ráp xác suất của các máy dò bộ phận mạnh mẽ. *ECCV lần thứ 8, Praha, Cộng hòa Séc*, tập I, trang 69–81, 2004.

[17] A. Mohan, C. Papageorgiou, và T. Poggio. Phát hiện đối tượng dựa trên ví dụ trong hình ảnh theo thành phần.*PAMI*, 23(4):349–361, tháng 4 năm 2001.

[18] C. Papageorgiou và T. Poggio. Một hệ thống có thể đào tạo để phát hiện đối tượng.*IJCV*, 38(1):15–33, 2000.

[19] R. Ronfard, C. Schmid, và B. Triggs. Học cách phân tích hình ảnh của mọi người.*ECCV lần thứ 7, Copenhagen, Đan Mạch*, tập IV, trang 700–714, 2002.

[20] Henry Schneiderman và Takeo Kanade. Phát hiện đối tượng bằng cách sử dụng số liệu thống kê của các bộ phận.*IJCV*, 56(3):151–177, 2004.

[21] Eric L. Schwartz. Ánh xạ không gian trong phép chiếu giác quan linh trưởng: cấu trúc phân tích và sự liên quan đến nhận thức.*Điều khiển học sinh học*, 25(4):181–194, 1977.

[22] P. Viola, MJ Jones và D. Snow. Phát hiện người đi bộ bằng cách sử dụng các mẫu chuyển động và ngoại hình.*ICCV lần thứ 9, Nice, Pháp*, tập 1, trang 734–741, 2003.