ISSN: 1859-2171
e-ISSN: 2615-9562
TNU Journal of Science and Technology 225(06): 115 – 121

MỘT KỸ THUẬT PHÁT HIỆN NGƯỜI ĐI BỘ DỰA TRÊN ĐẶC TRƯNG CHUYỂN ĐỘNG

Vũ Đức Thái[1], Dương Thị Nhung[1*], Ngô Đức Vĩnh[2], Phùng Thế Huân[1]
[1]Trường Đại học Công nghệ Thông tin và Truyền thông – ĐH Thái Nguyên
[2]Trường Đại học Công nghiệp Hà Nội

TÓM TẮT

Phát hiện người đi bộ là vấn đề quan trọng trong nhiều bài toán ứng dụng của lĩnh vực xử lý ảnh, ví dụ như giám sát giao thông, phát hiện đột nhập, xe tự hành… Trong bài báo này, chúng tôi trình bày một kỹ thuật phát hiện người đi bộ dựa trên đặc trưng Haar mở rộng, kết hợp với các bộ phân lớp yếu được thực hiện dựa trên thuật toán Adaboost để đưa ra quyết định. Các đặc trưng này được tính toán dựa trên yếu tố chuyển động bởi sự sai khác giữa các cặp ảnh theo thời gian. Kỹ thuật đã được thử nghiệm và chứng tỏ được sự hiệu quả trên cơ sở dữ liệu PETS 2001 và một số dữ liệu thu tại Trường Đại học Thông tin Truyền thông – Đại học Thái Nguyên.

Từ khóa: *Phát hiện người đi bộ; Haar; Haar-like; Haar wavelet; Adaboost…*

Ngày nhận bài: 02/3/2020; Ngày hoàn thiện: 05/5/2020; Ngày đăng: 11/5/2020

A TECHNIQUE FOR PEDESTRIAN DETECTION BASED ON MOTION FEATURES

Vu Duc Thai[1], Duong Thi Nhung[1*], Ngo Duc Vinh[2], Phung The Huan[1]
[1]TNU - University of Information and Communication Technology
[2]HaUI – Hanoi University of Industry

ABSTRACT

Pedestrian detection is an important issue in many application areas of image processing, such as traffic monitoring, intrusion detection, self-driving car... In this paper, we present a pedestrian detection technique based on extended Haar features combined with weak classifiers are implemented based on the Adaboost algorithm to make decisions. These features have been calculated based on the difference between pairs of images over time. The technique has been implemented and demonstrates the effectiveness on the 2001 PETS database.

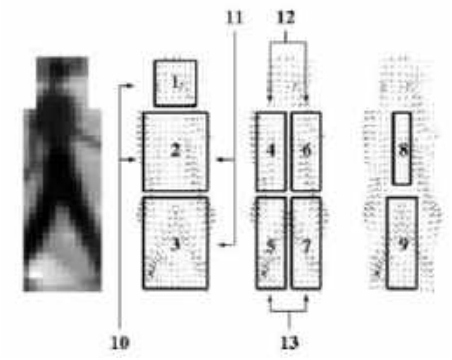
Keywords: *Pedestrian Detection; Haar; Haar-like; Haar wavelet; Adaboost…*

Received: 02/3/2020; Revised: 05/5/2020; Published: 11/5/2020

* Corresponding author. *Email: dtnhung@ictu.edu.vn*

1. Giới thiệu

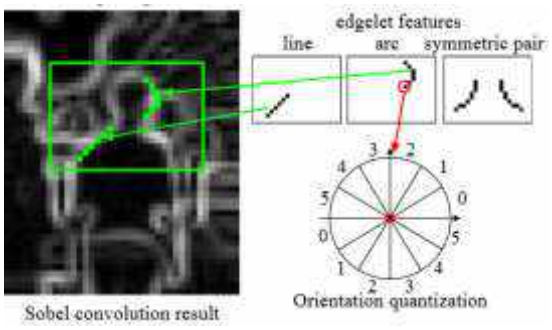
Bài toán phát hiện người đi bộ có thể được coi là một trường hợp riêng của bài toán phát hiện đối tượng. Một tiêu chí hay được nói đến trong phát hiện người đi bộ chính là quá trình đưa ra vết của người đi bộ từ các khung hình video. Quá trình này trọng tâm là quá trình xử lý chuỗi ảnh liên tiếp trong một đoạn video để phát hiện ra có hay không người đi bộ trong một đoạn hình ảnh.



Hình 1. *Các thành phần cục bộ với ảnh gradient* [1]

Đây là bài toán có nhiều thách thức và phức tạp do sự đa dạng trong diện mạo, tư thế, quần áo, màu sắc, cảnh nền… của người đi bộ. Ngoài ra điều kiện thời tiết, ánh sáng, khoảng cách quay, vấn đề che khuất… cũng ảnh hưởng đáng kể đến hiệu quả của việc phát hiện người đi bộ. Hiện có nhiều phương pháp, ý tưởng giải quyết bài toán này đã được nghiên cứu và đề xuất, mỗi phương pháp, ý tưởng có ưu điểm, nhược điểm riêng. Papageorgiou và Poggio [1] đã mô tả một hệ thống phát phát hiện người đi đường với tư cách là một phần của hệ thống hỗ trợ lái xe với khả năng biểu diễn đối tượng bằng việc sử dụng sự khác biệt cường độ, hướng trên nhiều mức giữa các vùng lân cận, và được tính toán với Haar wavelet; trên cơ sở đó, các đặc trưng được đưa vào mô hình máy vector hỗ trợ. Dalal và Triggs [2] thì xây dựng lược đồ các gradient có định hướng (HOG) để mô tả đối tượng. Theo đó, cửa sổ trượt sẽ được chia thành lưới các khối và các vector đặc trưng HOG sẽ được trích ra; sau đó đưa vào bộ phân lớp SVM tuyến tính. Kế thừa công trình này, Zhu và các đồng nghiệp [3] đẩy nhanh các tính năng HOG bằng cách sử dụng lược đồ histogram tích phân [4]. Shashua và các đồng nghiệp [5] đề xuất một biểu diễn tương tự đối với các thành phần cục bộ để xây dựng mô hình người (hình 1).

Với tiêu chí sử dụng các đặc trưng hình dạng, Gavrila và Philomin [6], [7] đã sử dụng khoảng cách Hausdorff và một hệ thống phân cấp mẫu để nhanh chóng kết hợp các biên ảnh vào một tập hợp các mẫu hình dạng. Wu và Nevatia [8] sử dụng một lượng lớn phân đoạn của các đoạn thẳng và đường cong ngắn, được gọi là các đặc trưng "edgelet", để biểu thị hình dạng cục bộ. Trong [9], "shapelets" là các bộ mô tả hình dạng được học phân biệt từ gradient trên các vùng cục bộ; tiếp cận boosting được sử dụng để kết hợp nhiều shapelets vào một bộ phát hiện tổng thể (hình 2). Ở kỹ thuật này, ban đầu, các đặc trưng cạnh được phát hiện bởi các kỹ thuật gradient được trích chọn trên các vùng cục bộ (hình 2 bên trái thể hiện kết quả với kỹ thuật gradient là Sobel), các đặc trưng này có thể là đoạn thẳng, cung, hoặc kết hợp với các vị trí và góc xoay khác nhau (hình 2 bên phải thể hiện các đặc trưng cạnh với vị trí và hướng khác nhau). Bước tiếp theo, một bộ phát hiện tổng thể theo tiếp cận boosting sẽ sử dụng kết hợp các đặc trưng này với nhau để đưa ra quyết định.



Hình 2. *Đặc trưng edgelet* [8]

Trong bài báo này, nhóm tác giả trình bày một kỹ thuật phát hiện người đi bộ dựa trên đặc trưng chuyển động, cụ thể là dựa trên sự sai khác giữa các cặp ảnh theo thời gian, và thông tin chuyển động được trích rút từ những sự sai

khác này. Phần tiếp theo của bài báo là cụ thể kỹ thuật phát hiện người đi bộ dựa trên đặc trưng chuyển động với một số vấn đề chi tiết hơn, đó là đặc trưng Haar mở rộng và kỹ thuật Adaboost. Phần 3 sẽ là thử nghiệm, đánh giá kết quả và cuối cùng là phần kết luận.

2. Phát hiện người đi bộ dựa trên đặc trưng chuyển động

2.1. Đặc trưng Haar mở rộng

Đặc trưng Haar mở rộng được đề xuất trong [10], được xây dựng dựa trên những đặc trưng Haar áp dụng trong bài toán phát hiện khuôn mặt trên ảnh. Những đặc trưng này được mở rộng để thực hiện trên sự sai khác giữa các cặp ảnh theo thời gian, và thông tin chuyển động có thể được trích rút từ những sự sai khác này. Ví dụ, vùng có tổng các giá trị tuyệt đối của các sự sai khác nếu lớn thì ứng với chuyển động. Thông tin về hướng chuyển động có thể được trích rút từ sự sai khác giữa các phiên bản đã dịch chuyển của ảnh thứ hai theo thời gian so với hình ảnh đầu tiên.

Các đặc trưng được áp dụng trên năm ảnh:

$$\Delta = \text{abs}(I_t - I_{t+1}) \quad (1)$$

$$U = \text{abs}(I_t - I_{t+1}\uparrow) \quad (2)$$

$$L = \text{abs}(I_t - I_{t+1}\leftarrow) \quad (3)$$

$$R = \text{abs}(I_t - I_{t+1}\rightarrow) \quad (4)$$

$$D = \text{abs}(I_t - I_{t+1}\downarrow) \quad (5)$$

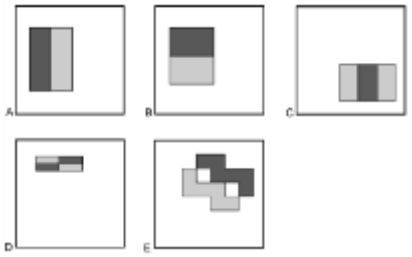
Với I_t và I_{t+1} là các ảnh theo thời gian, và $\{\uparrow, \downarrow, \leftarrow, \rightarrow\}$ là các toán tử dịch ảnh ($I_t\uparrow$ là I_tđã dịch lên trên bởi 1 pixel). Ví dụ như hình 3.

Hình 3. *Ví dụ về các ảnh đã dịch chuyển*

Một loại đặc trưng tính toán sự khác nhau giữa Δ và một trong {U, L, R, D}

$$f_i = r_i(\Delta) - r_i(S) \quad (6)$$

Với S là một trong {U, L, R, D} và $r_i(s)$ là một khung hình chữ nhật bên trong cửa sổ phát hiện. Các đặc trưng này trích rút thông tin về khả năng một vùng nào đó đang chuyển động theo một hướng nào đó (hình 4).



Hình 4. *Ví dụ đặc trưng Haar mở rộng áp dụng trên một ảnh*

Loại đặc trưng thứ hai so sánh tổng các vùng bên trong cùng một ảnh chuyển động:

$$f_j = \phi_j(S) \quad (7)$$

Với ϕ_j là một trong các đặc trưng được mô tả trong hình vẽ ở trên.

Cuối cùng, loại đặc trưng thứ ba đo cường độ của chuyển động từ một trong các ảnh chuyển động:

$$f_k = r_k(S) \quad (8)$$

Với S là một trong {U, L, R, D} và $r_k(s)$ là một khung hình chữ nhật bên trong cửa sổ.

Từ các đặc trưng, bộ phân lớp được xây dựng đơn giản là so sánh giá trị đặc trưng với một ngưỡng. Giá trị ngưỡng sẽ được học với từng bộ phân lớp cụ thể. Các bộ phân lớp này sẽ được kết hợp dựa trên kỹ thuật Adaboost.

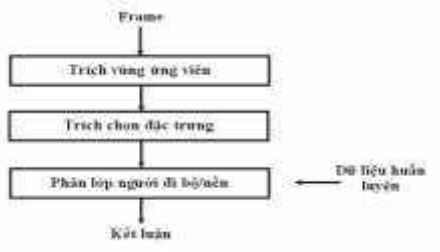
2.2. Adaboost

AdaBoost là một bộ phân loại mạnh phi tuyến phức dựa trên hướng tiếp cận boosting được Freund và Schapire đưa ra [11].

Hình 4. *Sơ đồ thuật toán Adaboost*

Adaboost hoạt động trên nguyên tắc kết hợp tuyến tính các bộ phân lớp yếu để hình thành một bộ phân lớp mạnh. Trong trường hợp này, các bộ phân lớp yếu chính là các bộ phân

lớp được tạo ra từ các đặc trưng Haar mở rộng đã được mô tả ở trên (chi tiết sơ đồ thuật toán theo hình 4).



Hình 5. *Sơ đồ tổng quát của hệ thống*

2.3. Quy trình hệ thống

Hệ thống được thực hiện dựa trên sơ đồ tổng quát như hình 5.

Bước trích vùng ứng viên sẽ lấy ra các vùng quan tâm từ ảnh để gửi đến khối trích chọn đặc trưng. Trong bước này nếu tránh được các vùng quan tâm không có người đi bộ càng nhiều thì tốc độ của hệ thống sẽ càng được cải thiện.Việc trích vùng ứng viên được thực hiện trong từng khung hình, cụ thể là dùng kỹ thuật cửa sổ trượt trên các vùng chuyển động của khung hình. Đầu tiên ta tính ảnh mặt nạ chuyển động. Ảnh mặt nạ chuyển động được tính thông qua kỹ thuật nền trung vị, cụ thể là khung hình hiện tại sẽ được so sánh với ảnh nền được tính bằng trung vị của *n* khung hình trước đó:

$$B(x,y,t) = median\{I(x,y,t-i)\},i=0,..,n-1 \quad (9)$$

Trong đó, $B(x,y,t)$ là giá trị điểm ảnh tại tọa độ (x,y) của nền tại thời điểm t, $I(x,y,t)$ là giá trị điểm ảnh tại tọa độ (x,y) của khung hình thu được tại thời điểm t. Việc tính ảnh mặt nạ chuyển động được thực hiện như sau:

$$|I(x,y,t)-B(x,y,t)\}| > threshold. \quad (10)$$

Như vậy, tại (x, y), nếu giá trị điểm ảnh hiện thời lệch so với nền vượt quá ngưỡng *threshold* thì (x,y) được gán nhãn là chuyển động. Sau đó, ta quét từng vùng khung hình có chuyển động bằng các cửa sổ có kích cỡ phù hợp để lấy ra các vùng ứng viên.

Bước trích đặc trưng chính là tính ra các giá trị đặc trưng Haar mở rộng trên vùng ứng viên đang xét. Để có thể tính toán một cách nhanh chóng, trước đó, sau khi nhận được khung hình hiện thời, ta thực hiện tính toán ảnh tích phân với các bước cụ thể sau:

- Từ khung hình hiện tại và khung hình trước đó xây dựng 5 ảnh Δ, U, D, L và R.
- Tính nhiều mức tỉ lệ (pyramids) các ảnh.
- Xây dựng các ảnh tích phân.

Ảnh tích phân là công cụ đã được Viola và đồng nghiệp [12] sử dụng để tính nhanh các đặc trưng Haar.

Bước cuối cùng là thực hiện phân lớp vùng ảnh ứng viên là người đi bộ hay nền. Việc phân lớp này được thực hiện dựa trên thuật toán Adaboost với các bộ phân lớp yếu sử dụng các đặc trưng chuyển động dựa trên Haar mở rộng.

3. Thử nghiệm

Chương trình được cài đặt bằng ngôn ngữ Matlab, sử dụng bộ công cụ Matlab R2015a. Matlabđược lựa chọn do khả năng đơn giản hóa việc giải quyết các bài toán tính toán kĩ thuật so với các ngôn ngữ lập trình truyền thống. Luồng thực hiện của chương trình tuân theo các bước của quy trình đã được mô tả. Việc thử nghiệm được tiến hành với hai trường hợp: trường hợp thứ nhất phương pháp sẽ được thử nghiệm với bộ dữ liệu PETS 2001 để kiểm chứng kết quả lý thuyết và trường hợp thứ hai chương trình sẽ chạy với một vài dữ liệu tự thu trong điều kiện thông thường tại Trường Đại học Thông tin Truyền thông – Đại học Thái Nguyên nhằm hướng đến đánh giá trong điều kiện video quay thực tế.

Bảng 1. *Dữ liệu huấn luyện và kiểm tra*

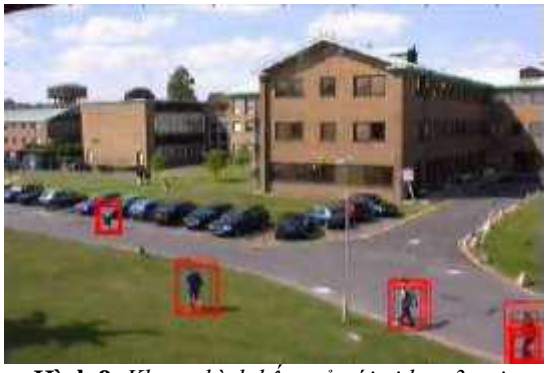
Dataset	Tập huấn luyện	Tập kiểm tra
1	video có 3063 frame	video có 2688 frame
2	video có 2989 frame	video có 2823 frame
3	video có 5563 frame	video có 5336 frame
4	video có 5010 frame	video có 6789 frame
5	video có 2866 frame	video có 2867 frame

Với trường hợp thứ nhất, dữ liệu video thử nghiệm được lấy từ cơ sở dữ liệu có sẵn PETS 2001 [13]. Đây là cơ sở dữ liệu gồm các ảnh và video quay người đi bộ thực hiện ngoài trời. Cơ sở dữ liệu này được xây dựng nhằm đánh giá hiệu quả của các thuật toán phát hiện người đi bộ. Đặc điểm của cơ sở dữ liệu này là dùng một camera để thu hình cảnh vật và người đi bộ. PETS 2001 gồm 5 tập dữ liệu, mỗi tập dữ liệu con có tập huấn luyện và kiểm tra tương ứng (bảng 1).

Tiến hành thử nghiệm với từng video dữ liệu, ta có với video_1.avi, khung hình có người đi bộ và xe đang di chuyển, kết quả phát hiện tương đối chính xác. Chương trình có khả năng phát hiện người đi bộ với kích thước nhỏ, ở khoảng cách xa (hình 6).

Hình 6. *Khung hình kết quả với video_1.avi: Nhiều người đi bộ và xe đang di chuyển*

Hình 7. *Khung hình kết quả với video_2.avi: Chỉ có xe đang di chuyển và người đi bộ bị che khuất bởi xe*



Hình 8. *Khung hình kết quả với video_3.avi: Nhiều người đi bộ đang di chuyển*

Với video_2.avi, khung hình có người đi bộ và xe đang di chuyển, kết quả phát hiện không được tốt (hình 7).

Video này cho thấy chương trình trong một số trường hợp với ảnh nền phức tạp hoặc có góc quay không được thuận lợi vẫn chưa phân biệt được người đi bộ với xe đang chuyển động. Để giải quyết vấn đề này cần đa dạng hóa tập dữ liệu huấn luyện với nhiều góc quay và nhiều khung cảnh khác nhau. Với video_3.avi, kết quả phát hiện khá tốt (hình 8).

Sau khi phát hiện ra vùng chuyển động, chương trình chỉ xem xét có phải là người đi bộ không nhưng do trong video này chỉ có người đi bộ chuyển động và không có các đối tượng chuyển động khác nên không có phát hiện nhầm (ví dụ với xe…).

Với video_4.avi, khung hình chỉ có người đi bộ đang di chuyển, kết quả phát hiện tương đối chính xác (hình 9).

Hình 9. *Khung hình kết quả với video_4.avi: Người đi bộ đang di chuyển trên đường và nền cỏ*

Sau khi thử nghiệm trên 4 video, khả năng phát hiện người đi bộ khoảng 80%. Trong một số trường hợp như ảnh nền phức tạp hoặc có góc quay không được thuận lợi vẫn chưa phân biệt được người đi bộ với các đối tượng khác đang chuyển động.

Trong trường hợp thứ hai, dữ liệu được tổ chức thu tại sân trường của Trường Đại học Công nghệ thông tin và truyền thông - Đại học Thái Nguyên. Dữ liệu được thu từ điện thoại di động, độ phân giải 1280x720, tốc độ 30 fps, thông số nén H264 - MPEG-4 AVC, bao gồm 4 video với thời gian quay là 2 phút 11 giây. Dữ liệu được thu với điều kiện đi lại bình thường của sinh viên cũng như các giảng viên trong sân trường (hình 10).

Hình 10. *Một số kết quả với dữ liệu thu tại Đại học Thông tin Truyền thông – Đại học Thái Nguyên*

Trong các kết quả thu được, ta nhận thấy rằng việc thực hiện phát hiện người đi bộ cho kết quả khá tốt trong những điều kiện đối tượng đứng riêng biệt, rõ ràng. Đây là cơ sở để có thể áp dụng thuật toán trong những ứng dụng có sử dụng video quay trong môi trường tự nhiên như điện thoại di động, camera giám sát. Bên cạnh đó, việc phát hiện cũng thỉnh thoảng bị nhầm với các đối tượng có đặc trưng cấu trúc trên ảnh tương tự như cây, góc xe ô tô. Ngoài ra, việc phát hiện cũng chưa được tốt trong những trường hợp đối tượng bị che khuất nhiều.

4. Kết luận

Người đi bộ là đối tượng được quan tâm trong nhiều hệ thống thị giác máy và phát hiện người đi bộ là vấn đề nghiên cứu cơ bản có nhiều tiềm năng ứng dụng thực tế.

Trong bài báo này, tác giả đã đề xuất một kỹ thuật phát hiện người đi bộ dựa trên sự sai khác giữa các cặp ảnh theo thời gian, với đặc trưng Haar mở rộng và kỹ thuật Adaboost. Kỹ thuật đã cài đặt thử nghiệm với cơ sở dữ liệu PETS 2001 và một số dữ liệu quay thực tế tại Trường Đại học Công nghệ Thông tin và Truyền thông - Đại học Thái Nguyên.

Tuy nhiên, kỹ thuật này mới chỉ tỏ ra có hiệu quả với các đối tượng đơn lẻ. Trong một số trường hợp như ảnh nền phức tạp hoặc có góc quay không được thuận lợi vẫn chưa phân biệt được người đi bộ với các đối tượng khác đang chuyển động.

Trong thời gian tới tác giả sẽ tiếp tục nghiên cứu cho những trường hợp đi theo đoàn và có sự che khuất, cũng như triển khai thử nghiệm trong các hệ thống video giám sát thực tế.

TÀI LIỆU THAM KHẢO/ REFERENCES

[1]. C. Papageorgiou, and T. Poggio, “A Trainable System for Object Detection,” *Int’l J. Computer Vision,* vol. 38, no. 1, pp. 15-33, 2000.

[2]. N. Dalal, and B. Triggs, “Histograms of Oriented Gradients for Human Detection,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2005, pp. 20-25.

[3]. Q. Zhu, S. Avidan, M. Yeh, and K. Cheng, “Fast Human Detection Using a Cascade of Histograms of Oriented Gradients,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2006, pp. 1491-1498.

[4]. F. M. Porikli, “Integral Histogram: A Fast Way to Extract Histograms in Cartesian Spaces,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2005, pp. 1-11.

[5]. Z. Shanshan et al., "Towards reaching human performance in pedestrian detection," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 973-986, 2017.

[6]. M. Jiayuan et al., "What can help pedestrian detection?" Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3127-3136.

[7]. D. M. Gavrila, “A Bayesian, Exemplar-Based Approach to Hierarchical Shape Matching,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 8, pp. 1408-1421, 2007.

[8]. B. Wu, and R. Nevatia, "Detection of Multiple, Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectors," Proc. 10th IEEE Int'l Conf. Computer Vision, 2005, pp. 90-97.

[9]. P. Sabzmeydani, and G. Mori, "Detecting Pedestrians by Learning Shapelet Features," Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2007, pp. 1093-1099.

[10]. P. A. Viola, M. J. Jones, and D. Snow, "Detecting Pedestrians Using Patterns of Motion and Appearance," *Int'l J. Computer Vision*, vol. 63, no. 2, pp. 153-161, 2005.

[11]. Y. Freund and R. E. Schapire, "A decision-theoretic generalization of online learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119-139, 1997.

[12]. V. Paul, and M. Jones, "Rapid object detection using a boosted cascade of simple features," Proceedings of the 2001 IEEE Computer Society Conference on, IEEE, 2001, vol. 1, pp. 511-518.

[13]. PETS, "Dataset," 2001. [Online]. Available: http://www.cvg.reading.ac.uk/PETS2001/pets2001-dataset.html. [Accessed Nov. 10, 2019].