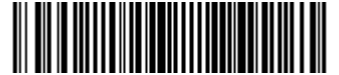




(12) **BẢN MÔ TẢ GIẢI PHÁP HỮU ÍCH THUỘC BẢNG ĐỘC QUYỀN
GIẢI PHÁP HỮU ÍCH**

(19) **Cộng hòa xã hội chủ nghĩa Việt Nam (VN)
CỤC SỞ HỮU TRÍ TUỆ**

(11)



2-0002480

(51)⁷ **G06K 9/00**

(13) **Y**

(21) 2-2016-00434

(22) 09/12/2016

(45) 25/11/2020 392

(43) 25/06/2018 363A

(73) Trường Đại học Công Nghệ Thông Tin (VN)

Khu phố 6, phường Linh Trung, quận Thủ Đức, thành phố Hồ Chí Minh

(72) Nguyễn Duy Khánh (VN); Dương Anh Đức (VN); Ngô Đức Thành (VN); Lê Đình
Duy (VN).

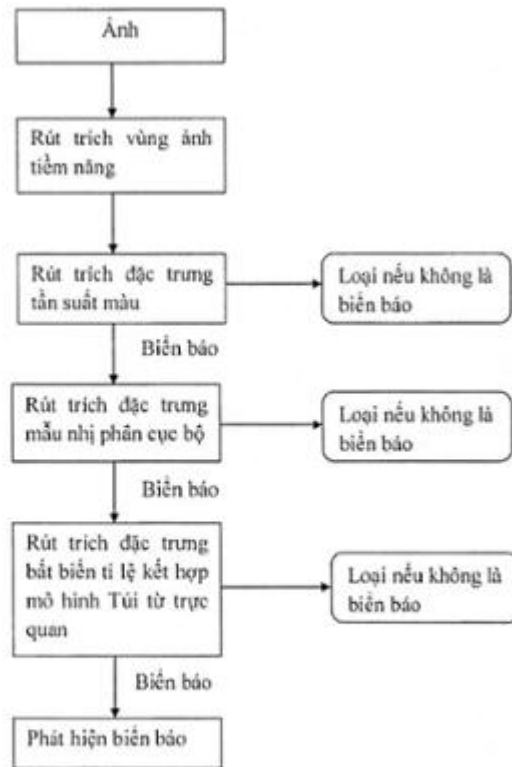
(54) **PHƯƠNG PHÁP PHÁT HIỆN BIỂN BÁO GIAO THÔNG SỬ DỤNG KẾT HỢP ĐA ĐẶC
TRUNG**

(57) Giải pháp hữu ích đề xuất phương pháp phát hiện biến báo giao thông sử dụng kết
hợp đa đặc trưng, phương pháp này bao gồm các bước:

i) rút trích vùng ảnh tiềm năng;

ii) rút trích đặc trưng ảnh (màu sắc, hoa văn, hình dáng, v.v..) bằng cách phân tầng đa
đặc trưng, bao gồm tầng đặc trưng tần suất màu (Hue, Saturation and Value-HSV), tầng
đặc trưng mẫu nhị phân cục bộ (Local Binary Pattern -LBP), và tầng đặc trưng bất biến tỉ
lệ (Scale Invariant Feature Transform - SIFT) kết hợp mô hình Túi từ trực quan (Bag of
Visual Words - BoVW);

iii) phát hiện biến báo bằng cách sử dụng thuật toán phân lớp máy học vector hỗ trợ
(SVM -Support Vector Machine).



Lĩnh vực kỹ thuật được đề cập

Giải pháp hữu ích thuộc lĩnh vực tin học, cụ thể giải pháp hữu ích đề xuất phương pháp phát hiện biển báo giao thông sử dụng kết hợp đa đặc trưng.

Tình trạng kỹ thuật của giải pháp hữu ích

Hiện nay số lượng biển báo giao thông đang được sử dụng là rất lớn, lắp đặt ở nhiều vị trí khác nhau, và thay đổi thường xuyên. Việc quản lý (theo dõi, bảo trì) các biển báo giao thông này thù công đòi hỏi rất nhiều nhân lực, chi phí. Vì vậy phương pháp phát hiện và nhận dạng biển báo giao thông tự động đã được triển khai trên các xe khảo sát chuyên dụng (gắn camera, thiết bị định vị GPS). Vị trí của biển báo, cũng như hiện trạng biển báo (mức độ hư hỏng) sẽ được thu thập tự động trong quá trình xe di chuyển. Dữ liệu thu thập được cũng có thể ứng dụng cho việc xây dựng bản đồ số, cho phép cập nhật các thông tin về hệ thống biển báo giao thông vào bản đồ một cách tự động. Từ đó phục vụ cho các ứng dụng khác như trong hệ thống dẫn đường, hệ thống giám sát, v.v.

Tuy nhiên, số lượng loại biển báo đa dạng và điều kiện ngoại cảnh phức tạp đặt ra nhiều thách thức để phương pháp phát hiện và nhận dạng tự động có thể hoạt động chính xác. Ngoài ra, để có thể áp dụng trong thực tế, vấn đề về mặt tốc độ của thuật toán trong phương pháp (với đòi hỏi kết quả chính xác cao) cũng rất quan trọng.

Nhiều kỹ thuật phát hiện đối tượng đã được áp dụng để giải quyết bài toán trong những năm gần đây. Điển hình là các công trình sau:

[1]. A. Mogelmose, M. M. Trivedi, T. B. Moeslund, Phát hiện và phân tích biển báo giao thông dựa trên thị giác dành cho các hệ thống hỗ trợ người lái thông minh: khảo sát và triển vọng (Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey). Kỷ yếu IEEE về các hệ thống giao thông thông minh năm 2012, trang 1484-1497 (IEEE Transactions on Intelligent Transportation Systems, 2012, p.p 1484–1497).

[2]. C. Bahlmann, Y. Zhu, V. Ramesh, M. Pellkofer, T. Koehler, Hệ thống phát hiện, theo vết, và nhận dạng biển báo giao thông sử dụng thông tin màu sắc, hình dáng và chuyển động (A system for traffic sign detection, tracking, and recognition using color, shape, and motion information), Kỷ yếu IEEE của hội nghị các phương tiện giao thông thông minh năm 2005, trang 255-260 (Intelligent Vehicles Symposium, 2005. Proceedings. IEEE, IEEE, 2005, p.p 255–260).

[3]. J. Levinson, J. Askeland, J. Becker, J. Dolson, D. Held, S. Kammel, J. Z. Kolter, D. Langer, O. Pink, V. Pratt, et al., Hướng tới lái xe hoàn toàn tự động: Hệ thống và các thuật toán (Towards fully autonomous driving: Systems and algorithms), Kỷ yếu IEEE của hội nghị các phương tiện giao thông thông minh năm 2011, trang 163-168 (Intelligent Vehicles Symposium, 2011 IEEE, p.p 163–168).

[4]. A. Gonz'alez, M. A. Garrido, D. F. Llorca, M. Gavil'an, J. P. Fern'andez, P. F. Alcantarilla, I. Parra, F. Herranz, L. M. Bergasa, M. 'A. Sotelo, et al., Hệ thống kiểm tra tự động biển báo giao thông và bảng thông báo sử dụng thị giác máy tính (Automatic traffic signs and panels inspection system using computer vision), Kỷ yếu IEEE về các hệ thống giao thông thông minh năm 2011, trang 485-499 (IEEE Transactions on Intelligent Transportation Systems, p.p 485–499).

[5]. W. Y. Yan, A. Shaker, S. Easa, Độ chính xác tiềm năng của các vị trí biển báo giao thông được rút trích từ ứng dụng xem đường phố của google (Potential accuracy of traffic signs' positions extracted from google street view), Kỷ yếu IEEE về các hệ thống giao thông thông minh năm 2013, trang 1011-1016 (IEEE Transactions on Intelligent Transportation Systems, 2013, p.p 1011–1016).

[6]. H. G'omez-Moreno, S. Maldonado-Basc'on, P. Gil-Jim'enez, S. Lafuente-Arroyo, Đánh giá về các thuật toán phân vùng cho nhận dạng biển báo giao thông (Goal evaluation of segmentation algorithms for traffic sign recognition), Kỷ yếu IEEE về các hệ thống giao thông thông minh năm 2010, trang 917-930 (IEEE Transactions on Intelligent Transportation Systems, 2010, p.p 917–930).

[7]. Z. Yankun, H. Chuyang, C. Wang, Hệ thống nhận dạng biển báo giới hạn tốc độ dạng tam giác hiệu quả thời gian thực (An efficient real time rectangle speed limit sign recognition system), Kỷ yếu IEEE của hội nghị các phương tiện giao thông thông minh năm 2010, trang 34-38 (Intelligent Vehicles Symposium (IV), 2010 IEEE, IEEE, 2010, p.p 34–38).

[8]. B. Hoferlin, K. Zimmermann, Hướng đến nhận dạng biển báo giao thông đáng tin cậy (Towards reliable traffic sign recognition), Kỷ yếu IEEE của hội nghị các phương tiện giao thông thông minh, năm 2009, trang 324-329 (Intelligent Vehicles Symposium, 2009 IEEE, IEEE, 2009, p.p 324–329).

[9]. X. Bar'ó, S. Escalera, J. Vitri' a, O. Pujol, P. Radeva, Nhận dạng biển báo giao thông sử dụng phương pháp phát hiện adaboost tiến hóa và phương pháp phân lớp forest-ecoc (Traffic sign recognition using evolutionary adaboost detection and forest-ecoc classification). Kỷ yếu IEEE về các hệ thống giao thông thông minh năm 2009, trang 113-126 (IEEE Transactions on Intelligent Transportation Systems, 2009, p.p 113–126).

[10]. Y. Xie, L.-f. Liu, C.-h. Li, Y.-y. Qu, Kết hợp đặc trưng thị giác nổi bật với học đặc trưng tần suất hướng cho bài toán phát hiện biển báo giao thông (Unifying visual saliency with hog feature learning for traffic sign detection). Kỷ yếu IEEE của hội nghị các phương tiện giao thông thông minh năm 2009, trang 24-29 (Intelligent Vehicles Symposium, 2009 IEEE, IEEE, 2009, p.p 24–29).

[11]. G. Overett, L. Petersson, Phát hiện biển báo giao thông qui mô lớn sử dụng các biến thể của đặc trưng tần suất hướng (Large scale sign detection using hog feature variants). Kỷ yếu IEEE của hội nghị các phương tiện giao thông thông minh năm 2011, trang 326-331 (Intelligent Vehicles Symposium, 2011 IEEE, IEEE, 2011, p.p 326–331).

[12]. G. Overett, L. Tychsen-Smith, L. Petersson, N. Pettersson, L. Andersson, Tạo các bộ phát hiện biển báo giao thông thông lượng cao bền vững sử dụng thống kê tần suất hướng trung tâm-bao quanh (Creating robust high-throughput traffic sign detectors using centre-surround hog statistics). Tạp chí thị giác máy và các ứng dụng năm 2014, trang 1-14 (Machine vision and applications, 2014, p.p 1–14).

[13]. S. Segvic, K. Brkić, Z. Kalafatić, A. Pinz, Khai thác các ràng buộc không gian và thời gian trong phát hiện biển báo giao thông từ phương tiện đang di chuyển (Exploiting temporal and spatial constraints in traffic sign detection from a moving vehicle), Tạp chí thị giác máy và các ứng dụng năm 2014, trang 1-17 (Machine vision and applications, 2014, p.p 1–17).

[14]. J. Greenhalgh, M. Mirmehdi, Phát hiện và nhận dạng biển báo giao thông trên đường theo thời gian thực (Real-time detection and recognition of road traffic signs), Kỷ yếu IEEE về các hệ thống giao thông thông minh năm 2012, trang 1498-1506 (IEEE Transactions on Intelligent Transportation Systems, 2012, p.p 1498–1506).

[15]. A. Vedaldi, V. Gulshan, M. Varma, A. Zisserman, Đa nhân cho phát hiện đối tượng (Multiple kernels for object detection), Kỷ yếu của hội nghị quốc tế IEEE về thị giác máy tính lần thứ 12 năm 2009, trang 606-613 (IEEE 12th International Conference on Computer Vision, IEEE, 2009, p.p 606–613).

[16]. B. Wu, R. Nevatia, Tối ưu hóa giữa tính phân biệt – hiệu năng trong việc tích hợp các đặc trưng cục bộ khác nhau cho phát hiện đối tượng (Optimizing discrimination-efficiency tradeoff in integrating heterogeneous local features for object detection), Kỷ yếu của hội nghị IEEE về thị giác máy tính và nhận dạng mẫu năm 2008, trang 1-8 (IEEE conference on Computer vision and pattern recognition, IEEE, 2008, p.p 1–8).

[17]. C. Wojek, S. Walk, B. Schiele, Phát hiện người bộ hành dùng thông tin đa nguồn (Multi-cue onboard pedestrian detection), Kỷ yếu của hội nghị IEEE về thị giác máy tính và nhận dạng mẫu năm 2009, trang 794-801 (IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2009, p.p 794–801).

[18]. J. Zhang, K. Huang, Y. Yu, T. Tan, Cấu trúc cục bộ tăng cường hog-lbp cho xác định vị trí đối tượng (Boosted local structured hog-lbp for object localization), Kỷ yếu của hội nghị IEEE về thị giác máy tính và nhận dạng mẫu năm 2011, trang 1393-1400 (IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2011, p.p 1393–1400).

[19]. G.Csurka, C.Dance, L.Fan, J.Willamowski, C.Bray. Phân loại thị giác với túi điểm then chốt (Visual categorization with bags of keypoints), Hội thảo về học thống kê trong thị giác máy tính (Workshop on statistical learning in computer vision), Kỷ yếu hội nghị châu Âu về thị giác máy tính năm 2004, trang 22 (European Conference on Computer Vision, 2004, p.p 22).

[20]. K.Mikolajczyk. C.Schmid, Đánh giá hiệu quả của các đặc trưng cục bộ (A performance evaluation of local descriptors), Kỷ yếu IEEE về phân tích mẫu và trí thông minh máy năm 2005, trang 1615-1630 (IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, p.p 1615–1630).

[21]. J.Zhang, M.Marszałek, S.Lazebnik, C.Schmid, Các đặc trưng cục bộ và nhân cho phân loại vân và chủng loại đối tượng: một nghiên cứu toàn diện (Local features and kernels for classification of texture and object categories: A comprehensive study), Tạp chí quốc tế về thị giác máy tính năm 2007, trang 213-238 (International journal of computer vision, 2007, p.p 213–238).

[22]. L.Zelnik-Manor. Về đặc trưng bất biến tỉ lệ và các tỉ lệ của chúng (On sifts and their scales), Kỷ yếu của hội nghị IEEE về thị giác máy tính và nhận dạng mẫu năm 2012, trang 1522–1528 (Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, p.p1522–1528).

[23]. A.Bosch, A.Zisserman, X.Muoz, Phân loại ảnh sử dụng rừng ngẫu nhiên (Image classification using random forests and ferns), Hội nghị quốc tế IEEE lần thứ 11 về thị giác máy tính năm 2007, trang 1–8 (IEEE 11th International Conference on Computer Vision, 2007, p.p 1-8).

Trong các phương pháp được đề xuất, đặc trưng dùng để biểu diễn biên báo có vai trò rất quan trọng. Các loại đặc trưng khác nhau đã được thử nghiệm, bao gồm phân phối màu (color distribution) [6], đặc trưng vân (như Local Binary Pattern – LBP [7], Haar-like [9, 13]), và đặc trưng hướng (gradient) (như Scale Invariant Feature Transform - SIFT [8], Histogram of Oriented Gradients -HOG [10, 11, 12, 14]). Tuy nhiên các loại đặc trưng trên đều tồn tại các khuyết điểm nhất định, ví dụ như phân phối màu sẽ kém

bền vững dưới điều kiện ánh sáng đa dạng, đặc trưng vẫn sẽ thiếu hiệu quả trong trường hợp đối tượng bị che khuất, trong khi đó đặc trưng hướng sẽ không đủ tính phân biệt dưới các biến đổi hình học bất thường.

Các nghiên cứu gần đây trong lĩnh vực phát hiện đối tượng tổng quát đã thử nghiệm rất thành công một mô hình biểu diễn mới, đó là mô hình Túi từ trực quan (Bag of Visual Words - BoVW). Mô hình này thường được kết hợp với với đặc trưng bất biến tỉ lệ (Scale Invariant Feature Transform - SIFT) [19, 20, 21].

Bên cạnh đó một hướng tiếp cận tiềm năng khác là kết hợp nhiều loại đặc trưng để tăng cường khả năng chống chịu và bổ khuyết cho nhau. Các nghiên cứu [15, 16, 17, 18] đã minh chứng hiệu quả vượt trội của phương pháp kết hợp này trên nhiều tập dữ liệu phổ biến. Tuy nhiên, sử dụng nhiều loại đặc trưng sẽ đòi hỏi chi phí tính toán rất lớn. Do đó để có thể áp dụng trong thực tế, cần thiết phải có các cải tiến để giảm thiểu chi phí tính toán.

Bản chất kỹ thuật của giải pháp hữu ích

Giải pháp hữu ích đề xuất phương pháp phát hiện biên báo giao thông sử dụng kết hợp đa đặc trưng với tốc độ xử lý nhanh, độ chính xác cao và chi phí tính toán thấp. Phương pháp phát hiện biên báo giao thông sử dụng kết hợp đa đặc trưng, phương pháp này bao gồm các bước:

- i) rút trích vùng ảnh tiềm năng;
- ii) rút trích đặc trưng ảnh (màu sắc, hoa văn, hình dáng, v.v..) bằng cách phân tầng đa đặc trưng, bao gồm tầng đặc trưng tần suất màu (Hue, Saturation and Value – HSV), tầng đặc trưng mẫu nhị phân cục bộ (Local Binary Pattern – LBP), và tầng đặc trưng bất biến tỉ lệ (Scale Invariant Feature Transform – SIFT) kết hợp mô hình Túi từ trực quan (Bag of Visual Words – BoVW):
- iii) phát hiện biên báo bằng cách sử dụng thuật toán phân lớp máy học vector hỗ trợ (SVM -Support Vector Machine):
 - trong đó khác biệt ở chỗ:
 - đặc trưng tần suất màu với vector biểu đồ tần suất có 7424 chiều;

đặc trưng mẫu nhị phân cục bộ với vector biểu đồ tần suất có 1711 chiều;
đặc trưng bất biến tỉ lệ kết hợp mô hình túi từ trực quan với vector biểu đồ tần suất
gồm 5000 chiều.

Mô tả vắn tắt hình vẽ

Hình 1 thể hiện lưu đồ của phương pháp phát hiện biển báo giao thông sử dụng kết hợp
đa đặc trưng.

Mô tả chi tiết giải pháp hữu ích

Phương pháp phát hiện biển báo giao thông sử dụng kết hợp đa đặc trưng bao gồm
các giai đoạn như sau :

Bước i) rút trích vùng ảnh tiềm năng trong tập dữ liệu chứa các hình ảnh biển báo.
Các hình ảnh này có thể chứa một hoặc nhiều biển báo, hoặc không chứa biển báo nào.
Phương pháp sử dụng mô hình boosted Haar cascade/Viola-Jones để rút trích các vùng
ảnh ứng viên với độ phù cao một cách nhanh chóng, các vùng này sau đó sẽ được xác
nhận lại bằng các bộ lọc đa đặc trưng ở các bước sau.

Mô hình boosted Haar cascade/Viola-Jones sử dụng cơ chế phân tầng, kết hợp với
đặc trưng Haar-like và bộ phân lớp yếu (cây quyết định), có chi phí tính toán thấp tại mỗi
tầng, cho phép phát hiện đối tượng nhanh chóng. Ảnh sẽ được quét theo cách tiếp cận vét
cạn sử dụng thuật toán cửa sổ trượt (sliding window) – sử dụng cửa sổ dịch chuyển qua
các vị trí khác nhau trong ảnh với nhiều kích thước tỉ lệ khác nhau. Các vùng ảnh trong
cửa sổ sẽ được phân lớp biển báo/không phải biển báo bằng các bộ phân lớp nhị phân sắp
xếp liên tiếp nhau (weak classifiers) theo kiểu phân tầng. Các vùng ảnh được phân lớp là
không phải đối tượng (ảnh nền) tại mỗi tầng sẽ được loại bỏ, ngược lại sẽ được tiếp tục
chuyển sang tầng tiếp theo. Phần lớn các phần ảnh nền có thể dễ dàng loại bỏ ngay từ các
tầng đầu tiên nên thuật toán có khả năng hoạt động rất nhanh chóng.

Thông thường, phương pháp Viola-Jones trả về một số lượng lớn các vùng ảnh
chồng lấp nhau. Khi đó các cảnh báo sai có thể được loại bỏ dựa vào một thuật toán tinh
giảm dựa vào tiêu chí số lượng vùng chồng lấp cần thiết ứng với mỗi biển báo. Tham số
này sẽ được thiết lập sao cho các vùng tiềm năng có thể được rút trích trong khi tỉ lệ cảnh
báo sai được giữ ở mức hợp lí. Cụ thể hơn, phương pháp sử dụng các thông số sau cho

quá trình huấn luyện: số lượng mẫu huấn luyện được phát sinh dựa trên tập huấn luyện (bao gồm tập ảnh biển báo và tập ảnh nền) ứng với mỗi loại biển báo là 10000; tỉ lệ phủ tối thiểu tại mỗi tầng (minimum hit rate) là 0,999; tỉ lệ phát hiện sai tối đa tại mỗi tầng (maximum false alarm) là 0,3; số lượng tầng (stage) tối đa của mô hình huấn luyện là 13.

Bước ii) rút trích đặc trưng ảnh (màu sắc, hoa văn, hình dáng, v.v..) từ tập các vùng ảnh tiềm năng ở bước i) bằng cách phân tầng đa đặc trưng, bao gồm tầng đặc trưng tần suất màu (Hue, Saturation and Value – HSV), tầng đặc trưng mẫu nhị phân cục bộ (Local Binary Pattern – LBP), và tầng đặc trưng bất biến tỉ lệ (Scale Invariant Feature Transform – SIFT) kết hợp mô hình Túi từ trực quan (Bag of Visual Words – BoVW). Theo một trường hợp cụ thể, tầng 1 sử dụng biểu đồ tần suất màu, tầng 2 sử dụng biểu đồ tần suất đặc trưng mẫu nhị phân cục bộ (Local Binary Pattern – LBP) và tầng cuối sử dụng biểu đồ tần suất đặc trưng bất biến tỉ lệ kết hợp mô hình Túi từ trực quan (Bag of Visual Words – BoVW).

Rút trích đặc trưng ảnh dựa trên tần suất màu cụ thể như sau:

Để biểu diễn đặc trưng màu, phương pháp cơ bản dựa trên tính toán biểu đồ tần suất (histogram) màu trong không gian tần suất màu (không gian này được biểu diễn bởi 3 thành phần: “hue” – sắc độ, “saturation” – độ bão hòa, và “value” – độ sáng) được sử dụng. Trong đó thành phần H (sắc độ) được sử dụng với 16 chiều, thành phần S (độ bão hòa) được sử dụng 4 chiều, và thành phần V (độ sáng) được sử dụng 4 chiều. Kết quả vector biểu đồ tần suất màu sẽ có 256 chiều. Để kết hợp thêm thông tin không gian (spatial information), biểu đồ tần suất màu sẽ được tính toán trên các lưới không gian (spatial grids) 2x2, 3x3, và 4x4. Các vector biểu đồ tần suất này được kết nối lại để xây dựng vector biểu diễn cuối cùng với 7.424 chiều sử dụng phương pháp nối vector thông thường (concatenation).

Rút trích đặc trưng ảnh dựa trên mẫu nhị phân cục bộ cụ thể như sau:

Phương pháp rút trích đặc trưng mẫu nhị phân cục bộ sử dụng $(P \cdot R) = (8, 2)$, kết quả là vector biểu đồ tần suất thu được sẽ gồm $2^8 = 256$ chiều. Để giảm số lượng chiều của vector đặc trưng và cung cấp khả năng bất biến với phép quay, phương pháp áp dụng

phương pháp cải tiến đặc trưng mẫu nhị phân cục bộ. Phương pháp này chia các mã nhị phân thành hai dạng: đồng nhất (uniform) và không đồng nhất (non-uniform). Một mẫu được gọi là đồng nhất khi chúng chứa nhiều nhất 2 lần chuyển đổi từ trạng thái 0 đến 1 và ngược lại khi chúng ta duyệt các bit theo chiều vòng tròn. Mỗi một kiểu mẫu đồng nhất sẽ được gán một nhãn phân biệt, và tất cả các mẫu không đồng nhất được gán chung một nhãn. Khi sử dụng (8, R) lân cận, sẽ có tổng cộng 256 mẫu nhị phân, trong đó gồm 58 mẫu đồng nhất và 1 mẫu không đồng nhất, theo đó sẽ có 59 nhãn. Kết quả là đặc trưng LBP được thu gọn chỉ còn 59 chiều.

Để mô tả thông tin không gian, giải pháp rút trích đặc trưng mẫu nhị phân cục bộ trên các lưới không gian (spatial grids) 2x2, 3x3 và 4x4, sau đó kết hợp các vector đặc trưng này lại để tạo thành vector biểu diễn cuối cùng với 1,711 chiều sử dụng phương pháp nối vector thông thường (concatenation).

Rút trích đặc trưng ảnh dựa trên đặc trưng bất biến tỉ lệ kết hợp mô hình túi từ trực quan. Mô hình Túi từ trực quan biểu diễn đối tượng dựa trên vector thống kê số lần xuất hiện của các “từ trực quan”. Các “từ” này là các đặc trưng cục bộ, được xác định bằng cách gom nhóm các đặc trưng cục bộ rút trích từ tập đối tượng huấn luyện, mỗi từ chính là trung tâm của mỗi nhóm này. Sau đó, đối tượng sẽ được biểu diễn bằng cách lượng từ hóa các đặc trưng cục bộ của nó theo các “từ” đã tìm được (được gán cho “từ” gần nhất), từ đó xây dựng thành vector tần suất xuất hiện của các “từ”. Mô hình túi từ trực quan được sử dụng để biểu diễn đặc trưng của biển báo ở tầng này.

Các bước thực hiện của việc rút trích đặc trưng ảnh dựa trên đặc trưng bất biến tỉ lệ kết hợp mô hình túi từ trực quan như sau:

Bước i) rút trích đặc trưng đa tỉ lệ trong không gian màu RGB (Red-Green-Blue) bằng phương pháp lấy mẫu cách đều (dense sampling); trong đó các tỉ lệ lấy mẫu được sử dụng là 4/6, 6/6, và 8/6, và khoảng cách giữa các điểm lấy mẫu là 2 điểm ảnh (pixel);

Bước ii) xây dựng bộ từ điển các từ trực quan bằng thuật toán gom cụm K-means++ để gom nhóm các đặc trưng rút trích được ở bước i); trong đó số lượng nhóm là 1.000;

Bước iii) xây dựng biểu đồ tần suất (xuất hiện) của các từ trực quan bằng cách tính vector tần suất gồm 1,000 chiều trên lưới không gian (spatial grid) 2x2, và kết hợp vector tần suất tại mỗi ô trên lưới không gian tạo thành vector tần suất cuối gồm 5,000 chiều;

Bước iv) biểu diễn biến báo dùng mô hình túi từ trực quan thông qua vector tần suất tính toán ở bước iii).

Sau khi thực hiện các bước rút trích đặc trưng ảnh, phương pháp tiến hành thực hiện Bước iii) phát hiện biến báo bằng cách sử dụng thuật toán phân lớp máy học vector hỗ trợ (SVM -Support Vector Machine). Cụ thể, đầu tiên phương pháp xây dựng thuật toán phân lớp máy vector hỗ trợ như sau:

Cho tập huấn luyện (x_i, y_i) , $i = 1 \dots N$, với $x_i \in R^d$ và $y_i \in \{-1, 1\}$. Cần tìm một phân lớp $f(x)$ sao cho: $f(x_i) \begin{cases} \geq 0 & y_i = +1 \\ < 0 & y_i = -1 \end{cases}$. Phân lớp tuyến tính SVM có dạng $f(x) = w^T x + b$, để

huấn luyện SVM (cụ thể là vector trọng số w) chúng ta cần giải bài toán tối ưu hóa trên w :

$$\min_{w \in R^d} \|w\|^2 + C \sum_{i=1}^N \max(0, 1 - y_i f(x_i))$$

Bài toán này luôn có duy nhất một cực tiểu có dạng: $w = \sum_{i=1}^N \alpha_i y_i x_i$ (có thể giải bằng phương pháp Karush Kuhn Tucker). Trong đó, mỗi mẫu huấn luyện x_i tương ứng với một hệ số Lagrange α_i , các mẫu có $\alpha_i > 0$ là các vector hỗ trợ (support vector), các mẫu còn lại sẽ có $\alpha_i = 0$.

Đối với tập dữ liệu không thể phân tách tuyến tính được, phương pháp hàm nhân (kernel) được sử dụng nhằm ánh xạ dữ liệu sang không gian có chiều lớn hơn mà ở đó có thể phân tách tuyến tính được. Một số hàm nhân thông dụng như: hàm nhân tuyến tính (Linear kernel): $K(u, v) = u^T v$; hàm nhân đa thức (Polynomial kernel): $K(u, v) = (\gamma u^T v + \text{coef} + 0)^d \dots$; hàm nhân cơ sở bán kính (RBF kernel): $K(u, v) = \exp(-\gamma \|u - v\|^2)$; Hàm nhân xichma (Sigmoid kernel): $K(u, v) = \tanh(\eta u^T v + \text{coef} + 0)$.

Sau khi xây dựng thành công thuật toán, phương pháp tiến hành phát hiện biển báo giao thông sử dụng thuật toán phân lớp máy học vector hỗ trợ gồm:

i) tạo tập dữ liệu huấn luyện bao gồm đặc trưng được rút trích của tập ảnh biển báo và tập ảnh nền được gán nhãn; trong đó tập đặc trưng được rút trích của tập ảnh biển báo được gán nhãn +1 và tập đặc trưng được rút trích của tập ảnh nền được gán nhãn -1;

ii) huấn luyện phân tầng cho mô hình máy học vector hỗ trợ dựa trên tập dữ liệu huấn luyện, trong đó mô hình máy học vector hỗ trợ bao gồm: tầng mô hình máy học vector hỗ trợ dùng đặc trưng tần suất màu, tầng mô hình máy học vector hỗ trợ dùng đặc trưng mẫu nhị phân cục bộ, và tầng mô hình máy học vector hỗ trợ dùng đặc trưng bất biến tỉ lệ kết hợp mô hình túi từ trực quan;

iii) phát hiện biển báo bằng cách chọn các ảnh phù hợp với cả ba tầng mô hình máy học vector hỗ trợ. Các vùng ảnh đã được rút trích đặc trưng ở bước ii) sẽ được đưa qua các bộ phân lớp theo thứ tự: tầng mô hình máy học vector hỗ trợ dùng đặc trưng tần suất màu, tầng mô hình máy học vector hỗ trợ dùng đặc trưng mẫu nhị phân cục bộ, và tầng mô hình máy học vector hỗ trợ dùng đặc trưng bất biến tỉ lệ kết hợp mô hình túi từ trực quan. Nếu thỏa mãn ở tầng trước (có nhãn trả về là +1/tức là biển báo) thì sẽ được chuyển tiếp sang tầng sau, ngược lại (có nhãn trả về là -1/tức là ảnh nền) thì sẽ bị loại ngay tại tầng này và không xử lý thêm nữa. Các vùng ảnh thỏa mãn cả 3 tầng được xem là kết quả phát hiện biển báo cuối cùng.

Ví dụ thực hiện giải pháp hữu ích

Ví dụ thực hiện giải pháp hữu ích trên tập dữ liệu GTSDDB (Germany Traffic Sign Detection Benchmark). Tập dữ liệu này gồm 600 ảnh chia thành 2 tập, tập huấn luyện gồm 300 ảnh và tập kiểm tra gồm 300 ảnh. Mỗi ảnh có thể chứa một hoặc nhiều biển báo, hoặc không chứa biển báo nào. Các biển báo được chia làm 3 loại: biển báo cấm, biển báo nguy hiểm, và biển báo chỉ dẫn.

Bước i) rút trích vùng ảnh tiềm năng trong tập dữ liệu ảnh gồm 600 ảnh ở trên. Phương pháp sử dụng mô hình boosted Haar cascade/Viola-Jones với cơ chế phân tầng.

kết hợp với đặc trưng Haar-like và bộ phân lớp yếu (cây quyết định) để quét tập ảnh huấn luyện hay kiểm tra theo cách tiếp cận vết cạn sử dụng thuật toán cửa sổ trượt (sliding window) – sử dụng cửa sổ dịch chuyển qua các vị trí khác nhau trong ảnh với nhiều kích thước tỉ lệ khác nhau. Các vùng ảnh trong cửa sổ sẽ được phân lớp biên báo/không phải biên báo bằng các bộ phân lớp nhị phân sắp xếp liên tiếp nhau (weak classifiers) theo kiểu phân tầng. Các vùng ảnh được phân lớp là không phải đối tượng (ảnh nền) tại mỗi tầng sẽ được loại bỏ, ngược lại sẽ được tiếp tục chuyển sang tầng tiếp theo. Phần lớn các phần ảnh nền có thể dễ dàng loại bỏ ngay từ các tầng đầu tiên nên thuật toán có khả năng hoạt động rất nhanh chóng.

Tiếp đó loại bỏ các cảnh báo sai bằng thuật toán tinh giảm dựa vào tiêu chí số lượng vùng chồng lấp cần thiết ứng với mỗi biên báo. Phương pháp sử dụng các thông số sau cho quá trình huấn luyện:

Số lượng mẫu huấn luyện được phát sinh dựa trên tập huấn luyện (bao gồm tập ảnh biên báo và tập ảnh nền) ứng với mỗi loại biên báo là 10000.

Tỉ lệ phủ tối thiểu tại mỗi tầng (minimum hit rate) là 0,999.

Tỉ lệ phát hiện sai tối đa tại mỗi tầng (maximum false alarm) là 0,3.

Số lượng tầng (stage) tối đa của mô hình huấn luyện là 13.

Bước ii) rút trích đặc trưng ảnh (màu sắc, hoa văn, hình dáng, v.v...) từ tập các vùng ảnh tiềm năng ở bước i) bằng cách phân tầng đa đặc trưng, bao gồm tầng đặc trưng tần suất màu (Hue, Saturation and Value – HSV), tầng đặc trưng mẫu nhị phân cục bộ (Local Binary Pattern – LBP), và tầng đặc trưng bất biến tỉ lệ (Scale Invariant Feature Transform – SIFT) kết hợp mô hình Túi từ trực quan (Bag of Visual Words – BoVW). Trong đó rút trích đặc trưng ảnh dựa trên biểu đồ tần suất màu thực hiện như sau:

Để biểu diễn đặc trưng màu, phương pháp tính toán biểu đồ tần suất (histogram) màu trong không gian tần suất màu (không gian này được biểu diễn bởi 3 thành phần: “hue” – sắc độ, “saturation” – độ bão hòa, và “value” – độ sáng) được sử dụng. Trong đó thành phần H (sắc độ) được sử dụng với 16 chiều, thành phần S (độ bão hòa) được sử

dụng 4 chiều, và thành phần V (độ sáng) được sử dụng 4 chiều. Kết quả vector biểu đồ tần suất màu sẽ có 256 chiều. Để kết hợp thêm thông tin không gian (spatial information), biểu đồ tần suất màu sẽ được tính toán trên các lưới không gian (spatial grids) 2x2, 3x3, và 4x4. Các vector biểu đồ tần suất này được kết nối lại để xây dựng vector biểu diễn cuối cùng với 7,424 chiều sử dụng phương pháp nối vector thông thường (concaternation).

Tiếp theo tiến hành rút trích đặc trưng ảnh dựa trên mẫu nhị phân cục bộ như sau:

Phương pháp rút trích đặc trưng mẫu nhị phân cục bộ sử dụng $(P, R) = (8, 2)$, kết quả là vector biểu đồ tần suất thu được sẽ gồm $2^8 = 256$ chiều. Để giảm số lượng chiều của vector đặc trưng và cung cấp khả năng bất biến với phép quay, phương pháp áp dụng phương pháp cải tiến đặc trưng mẫu nhị phân cục bộ. Phương pháp này chia các mã nhị phân thành hai dạng: đồng nhất (uniform) và không đồng nhất (non-uniform). Một mẫu được gọi là đồng nhất khi chúng chứa nhiều nhất 2 lần chuyển đổi từ trạng thái 0 đến 1 và ngược lại khi chúng ta duyệt các bit theo chiều vòng tròn. Mỗi một kiểu mẫu đồng nhất sẽ được gán một nhãn phân biệt, và tất cả các mẫu không đồng nhất được gán chung một nhãn. Khi sử dụng $(8, R)$ lân cận, sẽ có tổng cộng 256 mẫu nhị phân, trong đó gồm 58 mẫu đồng nhất và 1 mẫu không đồng nhất, theo đó sẽ có 59 nhãn. Kết quả là đặc trưng mẫu nhị phân cục bộ được thu gọn chỉ còn 59 chiều.

Để mô tả thông tin không gian, giải pháp rút trích đặc trưng mẫu nhị phân cục bộ trên các lưới không gian (spatial grids) 2x2, 3x3 và 4x4, sau đó kết hợp các vector đặc trưng này lại để tạo thành vector biểu diễn cuối cùng với 1,711 chiều sử dụng phương pháp nối vector thông thường (concaternation).

Sau cùng tiến hành rút trích đặc trưng ảnh dựa đặc trưng bất biến tỉ lệ kết hợp mô hình túi từ trực quan:

Bước i) rút trích đặc trưng đa tỉ lệ trong không gian màu RGB (Red-Green-Blue) bằng phương pháp lấy mẫu cách đều (dense sampling): trong đó các tỉ lệ lấy mẫu được sử dụng là 4/6, 6/6, và 8/6, và khoảng cách giữa các điểm lấy mẫu là 2 điểm ảnh (pixel):

Bước ii) xây dựng bộ từ điển các từ trực quan bằng thuật toán gom cụm K-means++ để gom nhóm các đặc trưng rút trích được ở bước i): trong đó số lượng nhóm là 1,000:

Bước iii) xây dựng biểu đồ tần suất (xuất hiện) của các từ trực quan bằng cách tính vector tần suất gồm 1,000 chiều trên lưới không gian (spatial grid) 2x2, và kết hợp vector tần suất tại mỗi ô trên lưới không gian tạo thành vector tần suất cuối gồm 5,000 chiều;

Bước iv) biểu diễn biến báo dùng mô hình túi từ trực quan thông qua vector tần suất tính toán ở bước iii).

Sau khi thực hiện các bước rút trích đặc trưng ảnh, phương pháp tiến hành thực hiện Bước iii) phát hiện biến báo bằng cách sử dụng thuật toán phân lớp máy học vector hỗ trợ (SVM -Support Vector Machine). Cụ thể, đầu tiên phương pháp xây dựng thuật toán phân lớp máy vector hỗ trợ như sau:

Cho tập huấn luyện (x_i, y_i) , $i = 1 \dots N$, với $x_i \in R^d$ và $y_i \in \{-1, 1\}$. Cần tìm một phân lớp

$f(x)$ sao cho: $f(x_i) = \begin{cases} \geq 0 & y_i = +1 \\ < 0 & y_i = -1 \end{cases}$. Phân lớp tuyến tính SVM có dạng $f(x) = w^T x + b$, để

huấn luyện SVM (cụ thể là vector trọng số w) chúng ta cần giải bài toán tối ưu hóa trên w :

$$\min_{w \in R^d} \|w\|^2 + C \sum_{i=1}^N \max(0, 1 - y_i f(x_i))$$

Bài toán này luôn có duy nhất một cực tiểu có dạng: $w = \sum_{i=1}^N \alpha_i y_i x_i$ (có thể giải bằng

phương pháp Karush Kuhn Tucker). Trong đó, mỗi mẫu huấn luyện x_i tương ứng với một hệ số Lagrange α_i , các mẫu có $\alpha_i > 0$ là các vector hỗ trợ (support vector), các mẫu còn lại sẽ có $\alpha_i = 0$.

Đối với tập dữ liệu không thể phân tách tuyến tính được, phương pháp hàm nhân (kernel) được sử dụng nhằm ánh xạ dữ liệu sang không gian có chiều lớn hơn mà ở đó có thể phân tách tuyến tính được. Một số hàm nhân thông dụng như: hàm nhân tuyến tính (Linear kernel): $K(u, v) = u^T v$;

hàm nhân đa thức (Polynomial kernel):

$K(u, v) = (\gamma u^T v + \text{coef} \cdot 0)^d \dots$; hàm nhân cơ sở bán kính (RBF kernel):

$K(u, v) = \exp(-\gamma \|u - v\|^2)$; Hàm nhân xichma (Sigmoid kernel):

$K(u, v) = \tanh(\eta u^T v + \text{coef} \cdot 0)$.

Sau khi xây dựng thành công thuật toán, phương pháp tiến hành phát hiện biến báo giao thông sử dụng thuật toán phân lớp máy học vector hỗ trợ gồm:

i) tạo tập dữ liệu huấn luyện bao gồm đặc trưng được rút trích của tập ảnh biến báo và tập ảnh nền được gán nhãn; trong đó tập đặc trưng được rút trích của tập ảnh biến báo được gán nhãn +1 và tập đặc trưng được rút trích của tập ảnh nền được gán nhãn -1;

ii) huấn luyện phân tầng cho mô hình máy học vector hỗ trợ dựa trên tập dữ liệu huấn luyện, trong đó mô hình máy học vector hỗ trợ bao gồm: tầng mô hình máy học vector hỗ trợ dùng đặc trưng tần suất màu, tầng mô hình máy học vector hỗ trợ dùng đặc trưng mẫu nhị phân cục bộ, và tầng mô hình máy học vector hỗ trợ dùng đặc trưng bất biến tỉ lệ kết hợp mô hình túi từ trực quan;

iii) phát hiện biến báo bằng cách chọn các ảnh phù hợp với cả ba tầng mô hình máy học vector hỗ trợ. Các vùng ảnh đã được rút trích đặc trưng ở bước ii) sẽ được đưa qua các bộ phân lớp theo thứ tự: tầng mô hình máy học vector hỗ trợ dùng đặc trưng tần suất màu, tầng mô hình máy học vector hỗ trợ dùng đặc trưng mẫu nhị phân cục bộ, và tầng mô hình máy học vector hỗ trợ dùng đặc trưng bất biến tỉ lệ kết hợp mô hình túi từ trực quan. Nếu thỏa mãn ở tầng trước (có nhãn trả về là +1/tức là biến báo) thì sẽ được chuyển tiếp sang tầng sau, ngược lại (có nhãn trả về là -1/tức là ảnh nền) thì sẽ bị loại ngay tại tầng này và không xử lý thêm nữa. Các vùng ảnh thỏa mãn cả 3 tầng được xem là kết quả phát hiện biến báo cuối cùng.

Phương pháp phát hiện biến báo giao thông sử dụng kết hợp đa đặc trưng được ứng dụng trong hệ thống phát hiện biến báo giao thông tự động, phục vụ công tác theo dõi, bảo trì biến báo sẽ tiết kiệm rất nhiều nhân lực, chi phí. Độ chính xác khi phát hiện biến báo đạt được từ 92%-98% tùy loại biến báo với tốc độ xử lý theo thời gian thực (~10 frames/second) ngoại trừ các trường hợp biến báo biến dạng quá lớn (do góc chụp camera, hư hỏng).

YÊU CẦU BẢO HỘ

1. Phương pháp phát hiện biển báo giao thông sử dụng kết hợp đa đặc trưng, phương pháp này bao gồm các bước:

- i) rút trích vùng ảnh tiềm năng;
- ii) rút trích đặc trưng ảnh (màu sắc, hoa văn, hình dáng, v.v..) bằng cách phân tầng đa đặc trưng, bao gồm tầng đặc trưng tần suất màu (Hue, Saturation and Value – HSV), tầng đặc trưng mẫu nhị phân cục bộ (Local Binary Pattern – LBP), và tầng đặc trưng bất biến tỉ lệ (Scale Invariant Feature Transform – SIFT) kết hợp mô hình Túi từ trực quan (Bag of Visual Words – BoVW);

iii) phát hiện biển báo bằng cách sử dụng thuật toán phân lớp máy học vectơ hỗ trợ (SVM -Support Vector Machine);

trong đó khác biệt ở chỗ:

đặc trưng tần suất màu với vectơ biểu đồ tần suất có 7424 chiều;

đặc trưng mẫu nhị phân cục bộ với vectơ biểu đồ tần suất có 1711 chiều;

đặc trưng bất biến tỉ lệ kết hợp mô hình túi từ trực quan với vectơ biểu đồ tần suất gồm 5000 chiều.

2. Phương pháp theo điểm 1, trong đó đặc trưng bất biến tỉ lệ kết hợp mô hình túi từ trực quan được rút trích theo quy trình bao gồm các bước:

i) rút trích đặc trưng đa tỉ lệ trong không gian màu RGB (Red-Green-Blue) bằng phương pháp lấy mẫu cách đều (dense sampling); trong đó các tỉ lệ lấy mẫu được sử dụng là 4/6, 6/6, và 8/6, và khoảng cách giữa các điểm lấy mẫu là 2 điểm ảnh (pixel);

ii) xây dựng bộ từ điển các từ trực quan bằng thuật toán gom cụm K-means++ để gom nhóm các đặc trưng rút trích được ở bước i); trong đó số lượng nhóm là 1.000;

iii) xây dựng biểu đồ tần suất (xuất hiện) của các từ trực quan bằng cách tính vectơ tần suất gồm 1.000 chiều trên lưới không gian (spatial grid) 2x2, và kết hợp vectơ tần suất tại mỗi ô trên lưới không gian tạo thành vectơ tần suất cuối gồm 5.000 chiều;

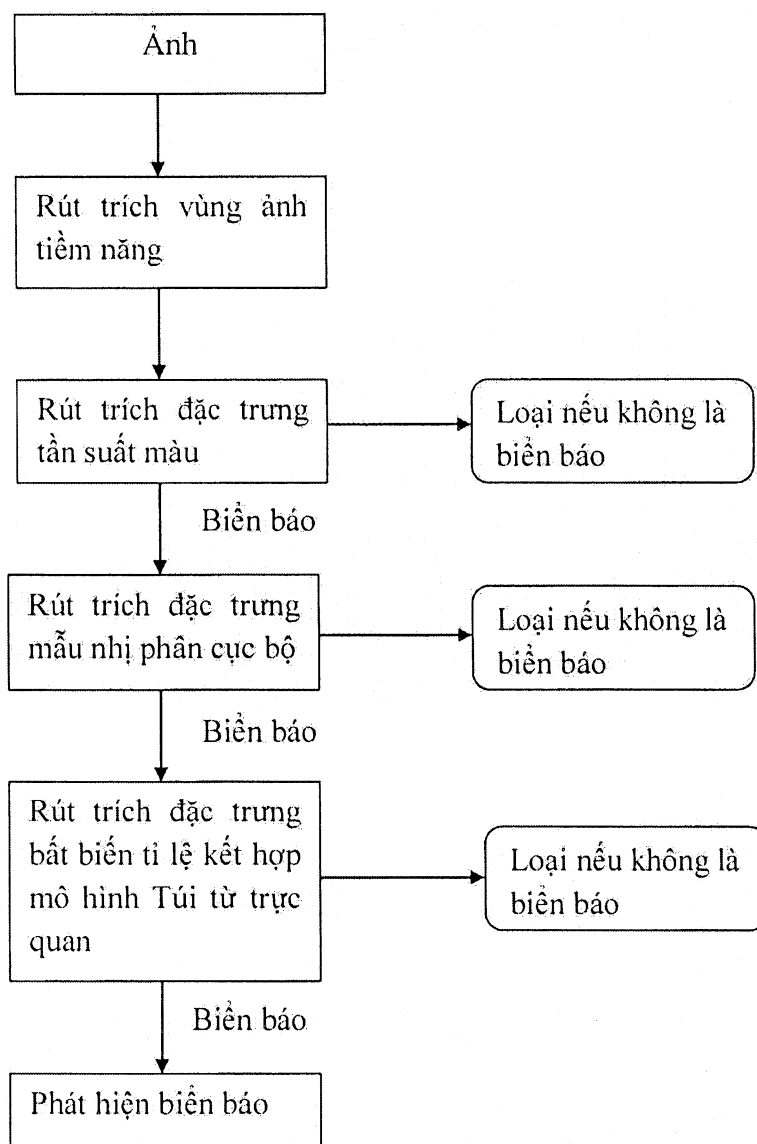
iv) biểu diễn biển báo dùng mô hình túi từ trực quan thông qua vectơ tần suất tính toán ở bước iii).

3. Phương pháp theo điểm 1, trong đó bước phát hiện biến báo giao thông sử dụng thuật toán phân lớp máy học vector hỗ trợ gồm các bước:

i) tạo tập dữ liệu huấn luyện bao gồm đặc trưng được rút trích của tập ảnh biến báo và tập ảnh nền được gán nhãn; trong đó tập đặc trưng được rút trích của tập ảnh biến báo được gán nhãn +1 và tập đặc trưng được rút trích của tập ảnh nền được gán nhãn -1;

ii) huấn luyện phân tầng cho mô hình máy học vector hỗ trợ dựa trên tập dữ liệu huấn luyện, trong đó mô hình máy học vector hỗ trợ bao gồm: tầng mô hình máy học vector hỗ trợ dùng đặc trưng tần suất màu, tầng mô hình máy học vector hỗ trợ dùng đặc trưng mẫu nhị phân cục bộ, và tầng mô hình máy học vector hỗ trợ dùng đặc trưng bất biến tỉ lệ kết hợp mô hình túi từ trực quan;

iii) phát hiện biến báo bằng cách chọn các ảnh phù hợp với cả ba tầng mô hình máy học vector hỗ trợ.



Hình 1