

NHẬN DẠNG VẬT BỊ BỎ LẠI BẰNG PHƯƠNG PHÁP HISTOGRAM SONG NỀN ỨNG DỤNG TRONG HỆ THỐNG GIÁM SÁT

Hoàng Văn Dũng

Trường Đại học Quảng Bình

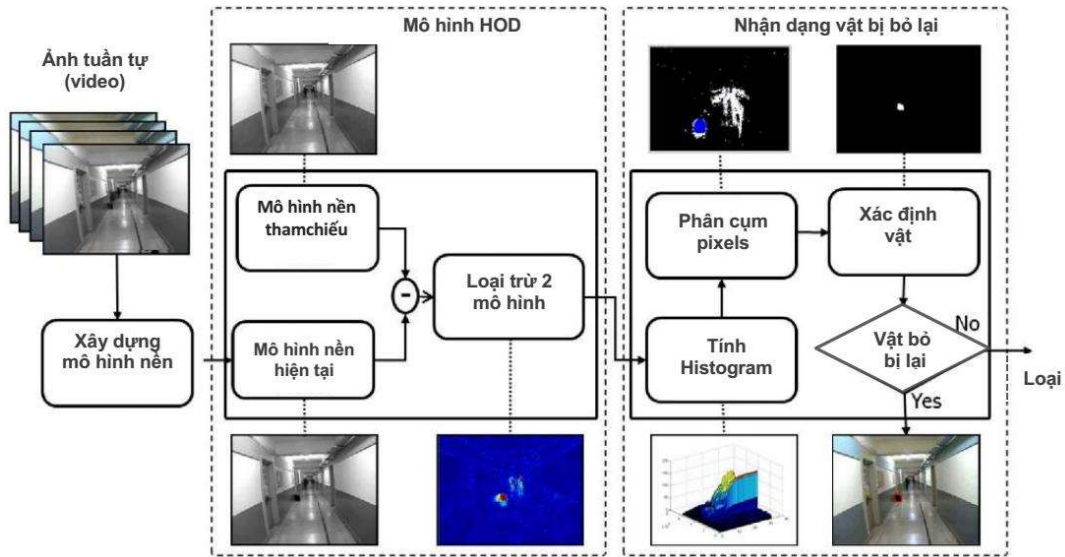
Tóm tắt. Cùng với sự phát triển nhanh chóng của khoa học và công nghệ, các hệ thống sử dụng kỹ thuật thị giác máy tính (computer vision) đã có nhiều ứng dụng và đạt được những kết quả nhất định, đặc biệt ứng dụng trong hệ thống giám sát công cộng. Trong bài toán giám sát công cộng, việc phát hiện đối tượng bị bỏ lại là nhiệm vụ hết sức quan trọng trong dự báo các sự cố có thể uy hiếp an ninh. Bài báo này giới thiệu phương pháp phát hiện vật bị bỏ lại dựa trên kỹ thuật Histogram song nền (Histogram of Dual Background -HOD) và kỹ thuật nhận dạng đối tượng. Phương pháp bao gồm các bước như sau: Xây dựng HOD theo điểm ảnh bằng cách tính độ khác biệt của giá trị điểm ảnh giữa ảnh hiện tại và các ảnh nền thu được trước đó. Các điểm ảnh bất biến được trích ra bằng kỹ thuật phân cụm trên mẫu HOD, các nhóm điểm ảnh được đánh giá dựa vào tính chất hình học và mối quan hệ với các đối tượng chuyển động để thu được các đối tượng tĩnh. Các đối tượng quan tâm sẽ được phân tích, phân loại bằng phương pháp nhận dạng đối tượng, ví dụ như SVM (support vector machine), Random forest, Neural network.... Kết quả đánh giá trên các cơ sở dữ liệu chuẩn cho thấy phương pháp này hiệu quả tốt và có khả năng ứng dụng trong các hệ thống giám sát an ninh.

Từ khóa: Nhận dạng mẫu, nhận dạng vật bị bỏ lại, hệ thống giám sát thông minh, học máy.

1. GIỚI THIỆU

Trong những năm qua, các hệ thống giám sát đã trở thành những ứng dụng quan trọng trong giám sát, đảm bảo an ninh công, hệ thống vận tải thông minh [1]. Các camera thường được gắn trên các ga tàu, trạm xe bus, sân bay, trường học và nhiều điểm công cộng khác. Mặc dù camera được lắp nhiều, có người theo dõi qua hệ thống, nhưng con người không thể bao quát và theo dõi thường xuyên tất cả camera để phát hiện những vấn đề bất thường. Do đó, cần thiết phải có những phần mềm hỗ trợ, phát hiện những bất thường để cảnh báo đến cán bộ giám sát, đảm bảo an ninh. Một trong những nhiệm vụ quan trọng là tự động phát hiện các đối tượng bị bỏ lại (túi mang cá nhân,...) để ngăn chặn những tình huống nguy hiểm xảy ra. Ví dụ, kẻ khủng bố đã đặt bom nơi công cộng tại Boston ngày 15 tháng 4 năm 2013, ba lô chứa bom được cố ý để lại tại nơi đang diễn ra cuộc thi marathon và gây nổ. Liên quan đến vấn đề này, nhóm tác giả [2] đã trình bày một số khái niệm liên quan đến đối tượng bị bỏ lại, xác định các đối tượng quan tâm và các ứng dụng của nó.

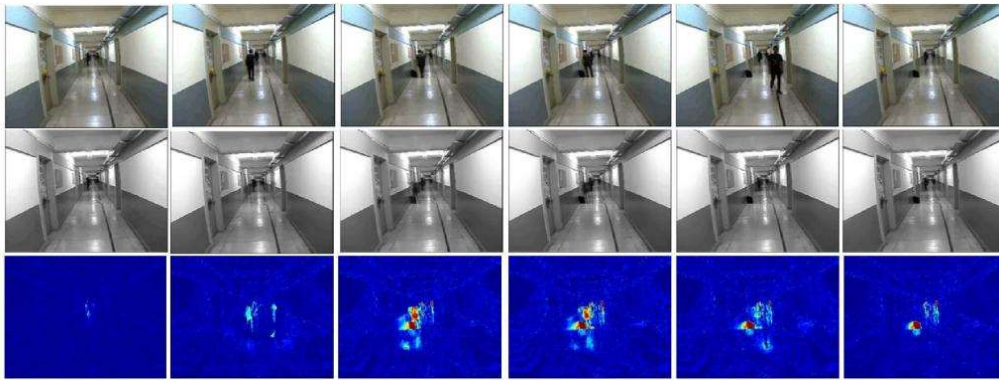
Trong lĩnh vực nhận dạng đối tượng đã có nhiều công trình nghiên cứu, thực nghiệm, điển hình như [3-11]. Trong phạm vi bài báo này, chúng tôi không tập trung vào nhận dạng đối tượng mà sử dụng nó như một công cụ đã có sẵn. Thay vào đó, chúng tôi sẽ trình bày một số kỹ thuật để phát hiện đối tượng bị bỏ lại, sau đó đối tượng sẽ được nhận dạng bằng cách sử dụng các kỹ thuật nhận dạng [12-15]. Nội dung phương pháp đề xuất được mô tả theo các bước như trong Hình 1.



Hình 1. Sơ đồ phương pháp phát hiện vật bị bỏ lại.

2. XÂY DỰNG MẪU NỀN DỰA VÀO HOD

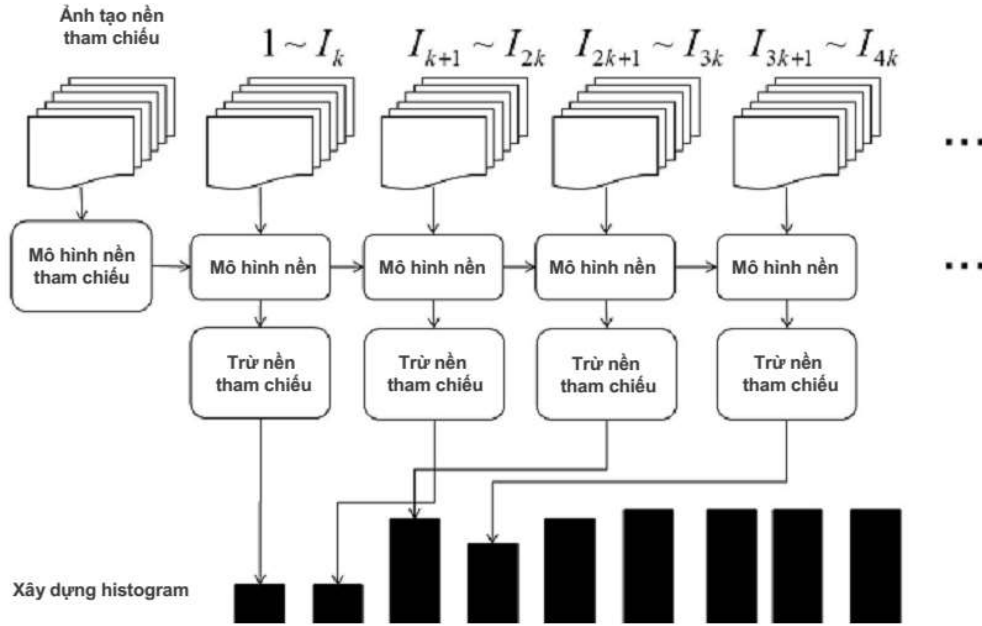
Phương pháp loại trừ nền (background subtraction-BG) truyền thống thường được dùng để tách vùng ảnh nền và vùng ảnh tiền cảnh. Loại trừ nền được thực hiện dựa vào việc so sánh mẫu nền với ảnh hiện tại, sau đó lọc cường độ điểm ảnh với ngưỡng xác định trước (hoặc ngưỡng được cập nhật động). Chi tiết về kỹ thuật một số phương pháp BG được trình bày trong [16],[17],[18]. Với phương pháp dựa trên HOD, thay vì tập trung vào việc tìm điểm ảnh tiền cảnh, hệ thống tập trung vào xác định điểm ảnh nền theo các mức độ khác nhau của mẫu nền hiện tại và mẫu nền tham chiếu. Công việc này được gọi là loại trừ hai mô hình nền, kết quả được minh họa như sau:



Hình 2. Loại trừ nền bằng HOD. Hàng thứ nhất là ảnh nền hiện tại, hàng thứ hai là mẫu nền tham chiếu, hàng cuối là giá trị trừ mẫu nền hiện tại với mẫu nền tham chiếu.

Với mỗi điểm ảnh, giá trị khác nhau $d_t(x,y)$ giữa mẫu nền tham chiếu $I_{B,R}(x,y)$ và mẫu nền hiện tại $I_{B,t}(x,y)$. Trong trường hợp giá trị $d_t(x,y)$ thay đổi đột ngột và sau đó nó trở lại mẫu nền, thì xác suất có đối tượng chuyển động trong vùng là rất cao. Các đối tượng tĩnh xuất hiện thì mẫu nền sẽ thay đổi từ từ cho đến khi mẫu nền bão hòa. Tập giá

trị $d_t(x,y)$ được kết hợp với nhau để xây dựng histogram trong một khoảng thời gian nhất định. Quá trình thực hiện được mô tả như Hình 3.



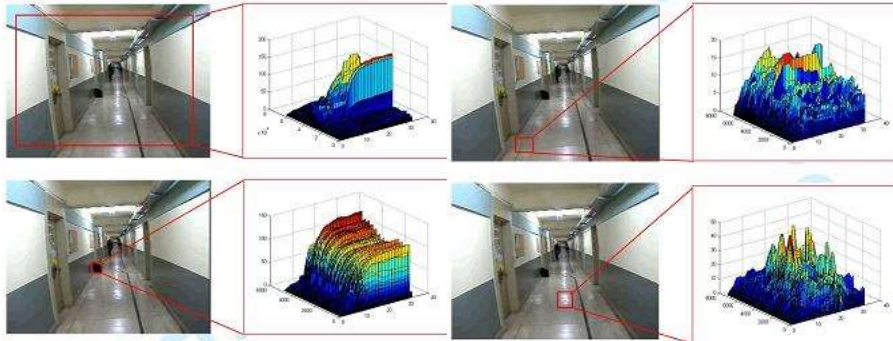
Hình 3. Quá trình tính giá trị HOD.

Kết quả HOD thể hiện sự thay đổi điểm ảnh theo thời gian liên quan đến mẫu nền, được tính như sau:

$$h(x, y) = \{d_{t-b+1}(x, y), \dots, d_{t-1}(x, y), d_t(x, y)\} \quad (1)$$

với b là số *bin* của histogram tương ứng với số các mẫu nền được sử dụng trong một khoảng thời gian nhất định.

Kết quả tính HOD được thể hiện trong Hình 4. Giá trị HOD của mỗi điểm ảnh được biểu diễn bằng vector 35 chiều.



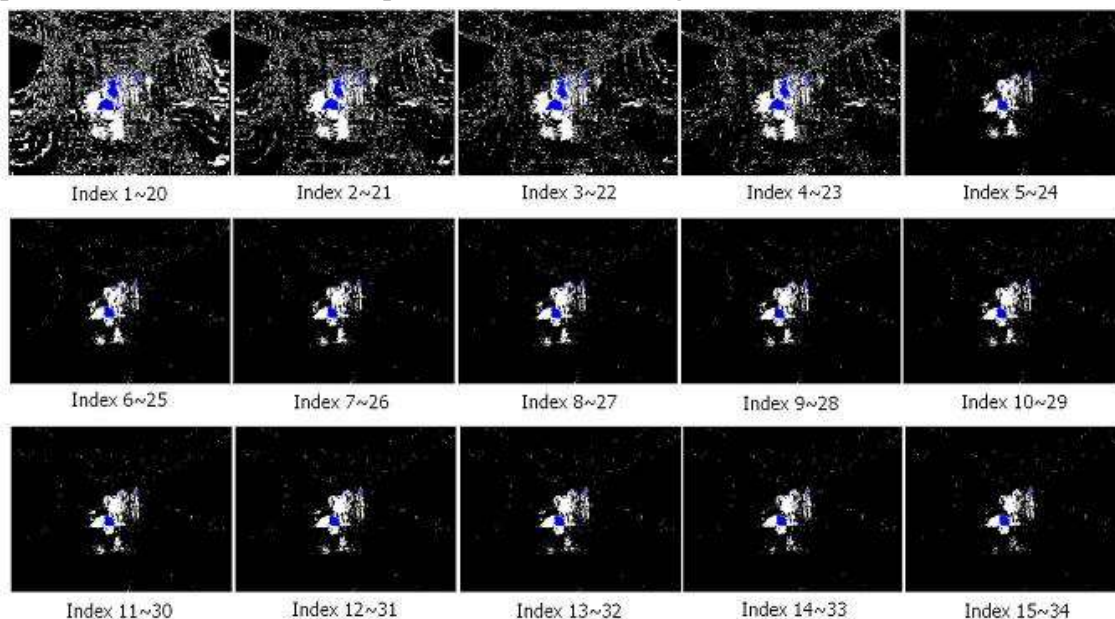
Hình 4. Giá trị HOD tương ứng với vùng đánh dấu đỏ, trong một khoảng 35 frame.

3. XÁC ĐỊNH TÚI ĐỒ BỊ BỎ LẠI

Để xác định túi đồ bị bỏ lại cần thực hiện các bước sau: dự đoán vùng ít biến đổi, sau đó xác định các đối tượng đứng yên, nhận dạng túi đồ và cuối cùng là dự đoán túi đồ đã bị bỏ lại.

Trước hết, các điểm ảnh được phân loại thành 3 kiểu: điểm ảnh nền (background), điểm ảnh nền trở nên không biến đổi (B2S: background –to- static) và điểm ảnh không

biến đổi sau trở thành ảnh nền (S2B: static – to – background). Trước hết, giá trị HOD sẽ được chuẩn hóa bằng chuẩn 2 (norm2). Giá trị histogram được sử dụng để so sánh trong phân nhóm điểm ảnh. Có nhiều phương pháp khác nhau có thể sử dụng để phân nhóm. Ở đây, phương pháp nhóm k -means được đề xuất sử dụng vì tính đơn giản và hiệu quả trong ứng dụng cụ thể. Trong ứng dụng này số cụm được sử dụng là 3. Ví dụ về phân nhóm điểm ảnh và kết quả được thể hiện trong Hình 5.



Hình 5. Phân nhóm các điểm ảnh: 3 nhóm điểm ảnh thể hiện ở các màu tương ứng: màu đen là background, màu xanh là B2S và màu trắng là S2B hoặc nhiễu.

Mối quan hệ giữa các vùng không biến động giữa ảnh hiện tại và ảnh trước đó được dùng để xóa nhiễu nhằm xác định đối tượng đứng yên một cách rõ ràng.

Sau khi xác định các đối tượng xuất hiện sau đó không biến đổi là giai đoạn nhận dạng đối tượng. Có nhiều phương pháp có thể sử dụng để nhận dạng túi đồ như dựa vào kích thước và tỷ lệ của đồ vật [19], hình dáng đồ vật [20] hoặc phương pháp học máy cũng có thể được sử dụng [13]. Bước cuối cùng là xác định túi đồ bị bỏ lại. Phương pháp trong [21] được dùng để xác định túi đồ bị bỏ lại dựa vào thời gian.

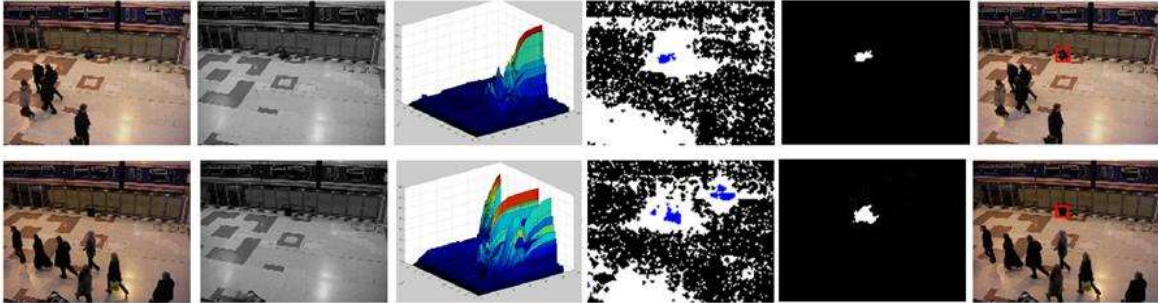


Hình 6. Dự đoán vùng có đối tượng đứng yên.

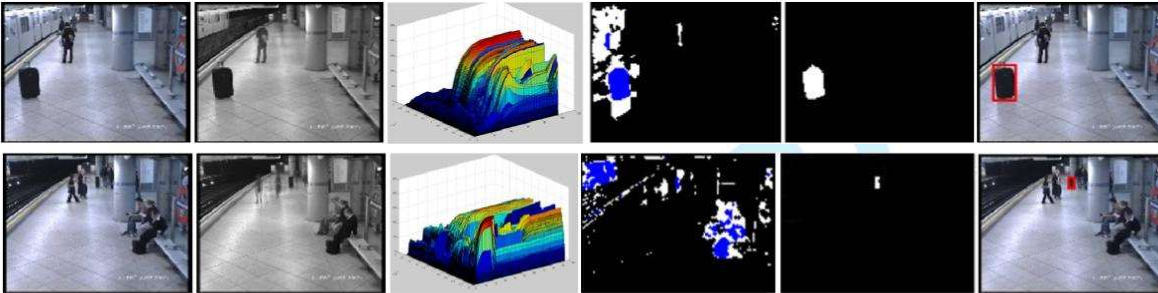
4. ĐÁNH GIÁ

Kết quả đánh giá thực hiện bằng ngôn ngữ C_{mex} (C++/Matlab), chip xử lý Core I7, RAM 8GB. Tham số được thiết lập bằng phương pháp “thử và sai” nhằm chọn giá trị tối ưu. Giá trị mẫu nền được cập nhật khi có khác biệt giữa giá trị lớn nhất và nhỏ nhất của cường độ sáng ảnh là bé hơn 30 ($\delta_f=30$), hệ số $\beta=0.3$, số bin histogram được

thiết lập là $b=30$. Vùng ảnh được xem là đứng yên ổn định với giá trị $\gamma=0.9$ và a_{\min} được xác định dựa vào độ phân giải ảnh. Tập dữ liệu huấn luyện gồm 200 mẫu (100 positive và 100 negative). Dữ liệu dùng để đánh giá được lấy từ PETS2006 [22] và AVSS2007[23], kết quả được thể hiện trong Hình 7 và Hình 8 tương ứng. Trong hình từ trái qua phải mô tả kết quả các bước như sau: Ảnh đầu vào, ảnh nền, giá trị HOD, phân cụm điểm ảnh, ước lượng vùng tĩnh có khả năng là đối tượng quan tâm, xác định vật bị bỏ lại.



Hình 7. Kết quả thực nghiệm trên dữ liệu PETS2006.



Hình 8. Kết quả thực nghiệm trên dữ liệu AVSS2007.

Sau đây là một số kết quả so sánh giữa phương pháp theo các tiêu chí đánh giá độ chính xác precision (P), recall (R) và F . Các tiêu chí này được tính theo các công thức:

$$P = tp / (tp + fp) \quad (2)$$

$$R = tp / (tp + fn) \quad (3)$$

$$F = 2 \times P \times R / (P + R) \quad (4)$$

với tp là số mẫu nhận dạng đúng positive, fp là số mẫu nhận dạng nhầm từ negative thành positive, fn là số mẫu nhận dạng nhầm từ positive thành negative.

Bảng 1. So sánh các phương pháp trên 2 cơ sở dữ liệu chuẩn

Phương pháp	PETS2006			AVSS2017		
	P	R	F	P	R	F
Lin2015 [24]	1.0	1.0	1.0	1.0	1.0	1.0
Chang 2013 [25]	1.0	1	1.0	1.0	1.0	1.0
Szwoch [20]	1.0	0.86	0.98	1.0	1.0	1.0
Maddalena 2013 [18]	-	-	-	1.0	1.0	1.0
Phương pháp đề xuất	1.0	1.0	1.0	1.0	1.0	1.0

Bảng 2. So sánh kết quả phân cụm

Kỹ thuật phân cụm	Thời gian (ms)	Đối tượng cố định			Vật bị bỏ lại		
		P	R	F	P	R	F
k -mean	32	0.85	1.0	0.91	1.0	1.0	1.0
GMM	121	0.76	1.0	0.80	1.0	1.0	1.0

5. KẾT LUẬN

Phương pháp trích vùng đối tượng khả nghi dùng kỹ thuật HOD là cách tiếp cận mới. HOD được tính toán dựa vào mức độ thay đổi vùng ảnh theo từng điểm ảnh trong một khoảng thời gian xác định để dự đoán vùng có đối tượng. Sau đó các vùng này được phân loại dựa vào phân tích biến đổi điểm ảnh theo giá trị histogram để tìm ra những vùng không biến đổi. Các đối tượng nghi vấn sẽ được phân loại dựa vào kỹ thuật nhận dạng đối tượng. Cuối cùng các đối tượng tĩnh được theo dõi trong một khoảng thời gian để xác định có phải đối tượng bị bỏ lại hay không. Phương pháp này đã được nghiên cứu và thực nghiệm trên một số cơ sở dữ liệu chuẩn và đạt được kết quả tốt. Một số hướng nghiên cứu tiếp cũng được xem xét như giải quyết bài toán trong trường hợp ánh sáng thay đổi nhanh chóng như có ánh đèn xe ban đêm, người qua lại đông đúc và đối tượng bị che khuất một phần,...

TÀI LIỆU THAM KHẢO

- [1] Seagate (2014), "Video surveillance trends report: An inside look at how it executives and system integrators are managing video surveillance data," *Technical Report*, pp. 1-15.
- [2] L. T. i (2006), "Imagery Library for Intelligent Detection Systems (i-LIDS); A Standard for Testing Video Based Detection Systems." pp. 75-80.
- [3] S. Munder, and D. M. Gavrilu (2006), "An Experimental Study on Pedestrian Classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1863-1868.
- [4] W.-S. Chen, P. C. Yuen, J. Huang, and D.-Q. Dai (2005), "Kernel machine-based one-parameter regularized Fisher discriminant method for face recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 35, no. 4, pp. 659-669.
- [5] N. Dalal, and B. Triggs (2005), "Histograms of oriented gradients for human detection." pp. 886-893.
- [6] X.-F. Wang, D.-S. Huang, and H. Xu (2010), "An efficient local Chan–Vese model for image segmentation," *Pattern Recognition*, vol. 43, no. 3, pp. 603-618.
- [7] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan (2010), "Object Detection with Discriminatively Trained Part-Based Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627-1645.
- [8] B. Li, C.-H. Zheng, and D.-S. Huang (2008), "Locally linear discriminant embedding: An efficient method for face recognition," *Pattern Recognition*, vol. 41, no. 12, pp. 3813-3821.

- [9] P. Dollar, C. Wojek, B. Schiele, and P. Perona (2012), "Pedestrian Detection: An Evaluation of the State of the Art," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 4, pp. 743-761.
- [10] P. Viola, M. J. Jones, and D. Snow (2003), "Detecting pedestrians using patterns of motion and appearance." pp. 734-741.
- [11] C. Papageorgiou, and T. Poggio (2000), "A Trainable System for Object Detection," *International Journal Compute Vision*, vol. 38, no. 1, pp. 15-33.
- [12] N. Cristianini, and J. Shawe-Taylor (2000), *An introduction to support vector machines and other kernel-based learning methods*: Cambridge university press.
- [13] C. Chih-Chung, and L. Chih-Jen (2011), "LIBSVM: a Library for Support Vector Machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1-27.
- [14] S. Maji, A. C. Berg, and J. Malik (2013), "Efficient Classification for Additive Kernel SVMs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 66-77.
- [15] C. C. Burges (1998), "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121-167, 1998/06/01.
- [16] C. Stauffer, and W. E. L. Grimson (1998), "Adaptive background mixture models for real-time tracking." pp. 1-252 Vol. 2.
- [17] Z. Zivkovic, and F. van der Heijden (2006), "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern recognition letters*, vol. 27, no. 7, pp. 773-780.
- [18] L. Maddalena, and A. Petrosino (2013), "Stopped object detection by learning foreground model in videos," *Neural Networks and Learning Systems, IEEE Transactions on*, vol. 24, no. 5, pp. 723-735.
- [19] N. K. P. T. T. Sebastian, and A. P. R. Collins (2006), "Multi-View Detection and Tracking of Travelers and Luggage in Mass Transit Environments," *PETS 2006*, pp. 67.
- [20] G. Szwoch (2016), "Extraction of stable foreground image regions for unattended luggage detection," *Multimedia Tools and Applications*, vol. 75, no. 2, pp. 761-786.
- [21] L. T. i (2006), "Imagery Library for Intelligent Detection Systems (i-LIDS); A Standard for Testing Video Based Detection Systems." pp. 75-80.
- [22] PETS2006BenchmarkData, "<http://www.cvg.reading.ac.uk/PETS2006/data.html>".
- [23] A. Cavallaro (2007), "<http://www.eecs.qmul.ac.uk/~andrea>".
- [24] K. Lin, S. C. Chen, C. S. Chen, D. T. Lin, and Y. P. Hung (2015), "Abandoned Object Detection via Temporal Consistency Modeling and Back-Tracing Verification for Visual Surveillance," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 7, pp. 1359-1370.
- [25] F. Porikli, Y. Ivanov, and T. Haga (2007), "Robust abandoned object detection using dual foregrounds," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, no. 1, pp. 1-11.

**AN ABANDONED OBJECT DETECTION FOR INTELLIGENT
SURVEILLANCE SYSTEM BASED ON HISTOGRAM OF DUAL
BACKGROUND SEGMENTATION**

***Abstract.** Nowadays, vision based surveillance systems are widely used for safety and security purposes in many fields. It is necessary for constructing an efficient intelligence system that enables analyzing and evaluating image content to recognize suspicious activities. One of the most important tasks is to recognize abandoned objects for unexpected incident prediction. The content of this paper consist of several tasks as follows: Per-pixel sequence of dual background is computed based on intensity difference between the current background model and the reference background model, which archived before within predefined period interval. Then, sequence patterns are clustered for extracting static pixels. These pixel regions are evaluated based on their relationship and geometry property to form moving object contours and extract regions of still objects. Finally, the interesting region results are further analyzed using object detector, e.g. SVM, Random forest, Neural network and so on. The effectiveness of the presented method is verified by implementing a framework for abandoned object segmentation and recognition on several standard dataset such as PETS2006, AVSS2007... It outperforms on recall, precision for abandoned object recognition. The results illustrated that the method is efficient and suitable to be applied to vision based surveillance systems.*

***Keywords:** Pattern recognition, abandoned object detection, intelligent surveillance system, machine learning.*