

# Empty Vehicle Detection with Video Analytics

Francesco Buemi<sup>1</sup>, Mariana Esposito<sup>2,3</sup>, Francesco Flammini<sup>2</sup>,  
Nicola Mazzocca<sup>3</sup>, Concetta Pragliola<sup>2</sup>, and Marcella Spirito<sup>1</sup>

<sup>1</sup> Aitek S.p.A, Via della Crocetta 15, Genoa Italy

{francesco.buemi,marcella.spirito}@aitек.it

<sup>2</sup> Ansaldo STS, Via Argine 425, Naples Italy

{francesco.flammini,concetta.pragliola}@ansaldo-sts.com

<sup>3</sup> University of Naples "Federico II", Via Claudio 21, Naples Italy  
{mariana.esposito,nicola.mazzocca}@unina.it

**Abstract.** An important issue to be addressed in transit security, in particular for driverless metro, is the assurance that a vehicle is empty before it returns to the depot. Customer specifications in recent tenders require that an automatic empty vehicle detector is provided. That improves system security since it prevents voluntary (e.g. in case of thieves or graffiti makers) or involuntary (e.g. in case of drunk or unconscious people) access of unauthorized people to the depot and possibly to other restricted areas. Without automatic systems, a manual inspection of the vehicle should be performed, requiring considerable personnel effort and being prone to failure. To address the issue, we have developed a reliable empty vehicle detection system using video content analytics techniques and standard on-board cameras. The system can automatically check whether the vehicles have been cleared from passengers, thus supporting the security staff and central control operators in providing a higher level of security.

**Keywords:** Transit security, empty vehicle detection, intelligent surveillance, video content analysis.

## 1 Introduction

Rail-based mass transit systems are vulnerable to criminal acts ranging from vandalism to terrorism. Physical security systems are aimed to protect all parts of the infrastructure, including tunnels, trains, platforms, public areas, technical rooms, depots, etc. The objective of those systems is to prevent and detect threats like aggressions and abnormal behaviours, vandalism, thefts, pick-pocketing, sabotage and terrorism [10]. In the last years, there has been an increased usage of video surveillance systems based on Video Content Analysis (VCA) [5]. The aim of VCA systems is to reveal potentially dangerous events through real-time analysis of video footage [9]. VCA is much less used on-board vehicles, due to several obstacles with respect to fixed installations, like constraints on camera view-field, high crowding, vibrations, light changes and other disturbs and limitations. However, those issues do not seem to apply when considering the empty train detection algorithm we address in this paper. In fact,

such a function is required to prevent access of unauthorized people (e.g. thieves or graffiti makers, drunk or unconscious people) to the depot and possibly to other restricted areas. For this function, either manual inspection of the vehicle or alternative automatic techniques seem to be expensive and error-prone. Among the alternative automatic techniques we cite passenger counting ([3] and [6]), crowded area estimation [11], infrared-based people detection ([2] and [7]). Those solutions require additional hardware and can be prone to failures due to occlusions and environmental conditions. The solution we describe and test in this paper is cost-effective, since it is based on already existing on-board hardware (network cameras and recorders), and it does not seem to suffer from any significant issues, as witnessed by the results of the tests we have performed on a real vehicle. The remainder of paper is structured as follow: Section 2 reports a short analysis of related literature; Section 3 describes the reference architecture for the system; Section 4 presents the algorithm used for empty vehicle detection; Section 5 provides an industrial case-study and test results; finally, Section 6 draws conclusions about performance and future improvements.

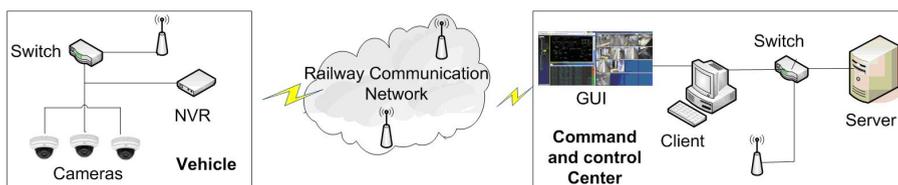
## 2 Related Works

The problem of detecting people/objects inside vehicles has been addressed in a broad range of applications with different techniques (e.g. infrared images, people counting, etc.) [3],[4], [5]. A similar solution based on video analytics has been described in [2]. The system has been developed by the University of Technology of Dresden and Bombardier Transportation: a video-based system is proposed to automatically check whether the people mover (driverless train for passenger transportation at airports) is empty, that is to say it does not contain any passengers and baggage. That solution works comparing the current image with a reference background image saved in appropriate conditions; the system is under development in order to be able to adapt to varying environmental conditions (e.g. sunlight, night) and to work on different assets, as will be described in the Section 4. With respect to that solution, the one we present in this paper is of more general application since it features several configuration parameters to adapt to any types of vehicle or even different assets (e.g. tunnel, depots, etc). The different configuration of parameters and working modes, including background comparison, allows to improve detection performance and to reduce false alarms. Therefore, the system allows the fine-tuning of the application to finally meet customer's requirements in terms of false negatives/positives with a limited effort.

## 3 Reference Architecture

Generally speaking, intelligent video analysis for mass-transit is provided by an image processing subsystem which is able to perform a real-time elaboration of

the footage coming from CCTV (Closed-Circuit TeleVision). Event detection is based on techniques which are able to identify and automatically tabulate (by comparison with the background) objects and individuals detected in the scene [12]. Event detection (e.g. unattended luggage, track crossing, etc.) is automatic; after an event, the system sends an alert to the control room operators, giving information about the type of event and displaying the related video streams. The reference video surveillance system for rail security is showed in figure 1. In each car two or more cameras are installed together with a Network Video Recorder (NVR) which is used to record on-board cameras and to run the video analytics algorithms. The NVR is provided with a wireless modem-router to connect to the ground access points and thus to allow real-time video surveillance on-board vehicles from the control rooms. A metropolitan area network connecting vehicles and control centers is hence required which can be based on Wi-Fi (ad-hoc infrastructures, used in metros) or public networks (WiMax, GSM/GPRS or UMTS, used in tramways). In general, depending on network bandwidth and signal strength, only the transmission of alarms or lower quality streams (low resolution, high compression ratio, low frame-rate, etc.) could be allowed. The control center is provided by client and video-wall for the operator, in which live stream and alarm can be collect from the vehicle. A possible variation to such architecture comes from running the video analytics algorithms not on the NVR but directly on the on-board (smart) cameras, thus shifting the computational load towards the periphery. In this case, in order to provide the required computing performance, cameras must meet minimum processing power specifications (e.g., Axis cameras with the ARTPEC-4 chipset) [8]. The automatic empty train detection system is based on events, managed by command and control center or by their subsystems.



**Fig. 1.** The reference video surveillance system architecture

## 4 The Algorithm for Empty Vehicle Detection

The Empty Vehicle detection algorithm is designed to detect the presence of objects and/or people in a monitored area inside a coach of a passenger train. The algorithm operates in two modes:

- A fully automatic mode, in which alarm notifications are generated if objects meeting certain requirements are detected in the scene;

- An asynchronous mode, driven by user request, in which alarm notifications are generated if a certain amount of change is detected between the current scene and a reference image.

More specifically, alarm notifications are generated in the first mode if the detected objects exceed a minimum dimension (*object\_min%*) or persons occupy more than a given percentage of the monitored area (*min%*), for more than a certain time (*time\_max*). The minimum dimension, minimum occupancy percentage and the maximum allowed permanence time can be set when configuring algorithm parameters. The second mode is triggered on user request and it is based on the difference between the image at the time of the request and a previously computed reference image: if such a difference is above a given percentage of the monitored area (namely, "Threshold Variation"), then an alarm is generated. Reference image creation is launched by the user (e.g. the possibly automatic train supervisor) and it should occur when the monitored area can be considered certainly empty (e.g. when the train is exiting the depot at the begin of its service). The minimum change percentage of the current scene compared to the reference image, as well as the duration of reference image creation, can be set during parameter configuration. In our application, the algorithm compares the images of the internal of the train with a reference image recorded after the train has left the depot at the beginning of its trip. If the difference exceeds the "Threshold Variation", an alarm is generated (figure 2). The operator can hence confirm the alarm by visual inspection using on-board CCTV cameras and, if necessary, stop the train and ask the stewards to go for a check.

Please note that the approach is not only based on background subtraction. In fact, for both modes, the analysis is not based only on the comparison of two images. The comparison and elaboration is based on the transformation of the captured image in order to extract several information useful for the recognition of non-empty vehicles. This transformation and information extraction provides higher robustness, in particular with respect to different lighting conditions, as will be shown in the Section 5. Furthermore, in the asynchronous mode, the reference for the comparison is not only the reference image (saved at the beginning), but a collection of images derived from the original image. In particular, reference images are produced for a configurable duration and stored for the comparison at the time the user requests the check.

## 5 Case-Study

The experimentation of the algorithm has been performed on a real vehicle for urban transportation. In our scenario, the empty vehicle detection algorithm is able to automatically check if the vehicle is empty before entering the depot at the end of its service (last station). Please note that the algorithm is not running during the rest of passenger service. In order to increase the reliability of the algorithm, both the modes described in the previous Section are enabled. For the asynchronous mode, the reference image is saved in the most appropriate



**Fig. 2.** Prototypical non-empty vehicle alarm

condition, that is at a specific time when the vehicle is leaving the depot before beginning passenger service, so that it can adapt to possible changes in the train internal (e.g. change of furniture, advertisement panels). After the algorithm is activated by an appropriate event, a counter is started to automatically stop the algorithm after  $n$  seconds. During this time, the algorithm compares the current images of the internal of the train with the reference images (asynchronous mode) to check whether the difference in terms of percent of pixel is more than the predefined "Threshold Variation". At the same time, the algorithm checks if there are objects exceeding the  $object\_min\%$  dimension or if a person occupies more than the preset percentage of the monitored area ( $min\%$ ), for more than the preset time ( $time\_max$ ) (automatic mode). The activation of both conditions is necessary in order to detect immobile persons (e.g. in case of drunk, unconscious or sleeping people) and to reduce false positives. In fact, the automatic mode is able to reliably detect small objects left in the scene or larger moving objects, while the asynchronous mode (i.e. background comparison) is able to reliably detect relevant variations in the scene (which can correspond to immobile persons). However, since in the asynchronous mode there is no possibility to configure the size of objects, in that mode it could easily happen that unrelated small variations in the scene (e.g. corresponding to distant areas) are summed up to produce a percentage variation which is higher than the threshold set for object detection, hence producing a false positive. Therefore, being able to independently set the parameters of the two modes allows to more accurately fine-tune the system to get a balanced performance in terms of false negative/positive.

## 5.1 Test Environment

The hardware consists of a network camera connected to a NVR running the VCA software. Events are simulated in different environmental conditions



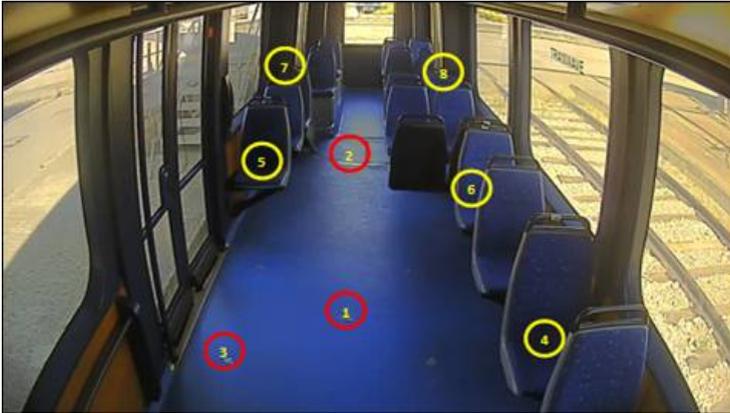
**Fig. 3.** Prototypical non-empty vehicle alarm

(including light variations, occlusions, etc.), ranging from optimal to worst cases. The first version of the algorithm we tested featured unacceptable performance. It was based on a combination of different algorithms but the analysis of the background was based only on a percent pixels variation and it did not work well in all conditions. Hence the algorithm has been improved using the new implementation described in this paper and the results showed a very good performance. While the first version of the algorithms was sensible to light variations (it worked well only with artificial lighting), the new one features good performance in any light and weather conditions, including possible varying shadows. The new version is also able to discriminate between objects and persons. In other words, the algorithm can give different alarms for persons and left luggage. For instance, we have simulated the presence of a sleeping person (that may be confused with an object) and the algorithm, with the right configuration, was able to correctly detect a person instead of an object. Of course, there may be problems in presence of obstacles (e.g. seats): to work best in discriminating between persons and objects, the algorithm should not be applied to areas where persons can be 'broken' in parts due to obstacles.

## 5.2 Test Results in the Asynchronous Mode

In this section, we report the results of a test campaign performed in the asynchronous mode for person detection, in which the main configuration parameter is "Threshold variation" and natural light is used. The other test conditions are:

- Area of background elaboration: this area includes the whole internal of vehicle, except the windows which create some problems due to the outside scene (Figure 3).
- Threshold variation= 10%, 15%, 20%.



**Fig. 4.** Test points for the empty vehicle algorithm

A person stands in the scene wearing dark clothes (the most critical condition since the background is dark), in eight different positions, covering fairly well all the peculiarities of the scene (Figure 4). In points 1, 2 and 3, the person is standing up; in the other positions, the person is seated. The person stays in each position for 30 seconds. In Table 1, we report the results of this test, where: TP = *True Positive*, TN = *True Negative*, FN = *False Negative*, FP = *False Positive*. The columns are associated to the positions, while the rows indicate the threshold variation set. The last column indicates the result of the algorithm when the person leaves the scene. The highest probability of detection is reached with a threshold of 10%. However, with such a value (and even increasing the threshold to 15%), the system gives a false positive alarm when the person leaves the scene, due to slight light variations in the background. The threshold of 20% seems to represent a good trade-off, since it is less sensible to slight variations and it only misses points 7 and 8, which are the most critical due to the presence of seats as obstacles to the field of view.

### 5.3 Test Results of Automatic Mode

In this section, we report results of a test campaign performed in the automatic mode, in which the main configuration parameters are:

- Change Detection Area: this area include the whole internal of the vehicle (same of Figure 3).
- Min percentage of occupied area (min%): 5%, 10%, 15%.
- Occupation time: 3 s.
- Min object dimension: 0.90%.

**Table 1.** Experimentation results in the Asynchronous mode

Position									
Threshold	1	2	3	4	5	6	7	8	Non Empty
10%	TP	FP							
15%	TP	TP	TP	TP	TP	TP	FN	FN	FP
20%	TP	TP	TP	TP	TP	TP	FN	FN	TN

Event conditions are the same of the test for the asynchronous mode (see previous section), including the eight reference positions. Furthermore, in each position, a bag is left in order to test object detection. If the person is not immobile, the algorithm correctly detects a person in all 8 positions with any min% value (i.e. 5%, 10%, 15%), and that works also in the positions 7 and 8, which represented criticalities for the Asynchronous mode, and even near the back window; however, if the person is immobile, this mode produces false negatives. Regarding the left objects, they were correctly detected only in the positions 1, 2, 3, 4 and 5, while in the others there were issues due to the presence of obstacles (other seats).

## 6 Conclusions

In this paper, we have described an automatic empty vehicle detection system tested on a real vehicle. Though the system was designed specifically for passenger trains, the system may also be used on different type of vehicles (e.g. freight cars) or sites (e.g. stations, tunnels), provided that appropriate tests are conducted in order to measure performance in presence of specific disturbs (e.g. ambient light, wind, rubbish, etc.). Therefore, the system is applicable to the protection of any assets within an infrastructure where there is the need for ensuring no people or objects are present during closing times. In order to significantly improve system performance, wherever possible, one should apply the following:

1. "Architectural design for security": vehicle internal should be designed such to avoid occlusions in camera views, e.g. using lateral seats with no under-seat space (this is important also for 'traditional' CCTV surveillance for the detection of left luggage).
2. Exclusions of areas with obstacles from one camera view and coverage of those areas using an additional camera (e.g. installed in the opposite side of the coach).

Whenever extremely high reliability is required, it is possible to integrate the video-based solution with further multi-technology sensors (e.g. passive IR, microwave, thermal view, etc.). We are going to test the system in "extreme" operational conditions (simulating the effect of stormy weather, lamp faults, vibrations, etc.) and studying the effects of variations of all the relevant algorithm parameters (which could possibly automatically adapted to the environment) and event characteristics (size, colour and speed of objects).

## References

1. Schuette, J., Scholz, S., Crawford, C.: An Automatic Empty Vehicle Detection System for Airport People Movers Applications. In: IEEE 75th Vehicular Technology Conference, VTC Spring (2012)
2. Bertozzi, M., Broggi, A., Graf, T., Grisleri, P., Meinecke, M.: Pedestrian detection in in-framed images. In: Proc. IEEE Intelligent Vehicles Symp., Columbus, OH, pp. 662–667 (June 2003)
3. Wusheng, W., Mengfen, H., Chunglin, H.: People tracking and counting for applications in video surveillance system. In: International Conference on Audio, Language and Image Processing. IEEE (July 2008)
4. Kyriakidis, M., Hirsch, R., Majumdar, A.: Metro railway safety: An analysis of accident precursors. *Safety Science* 50(7), 1535–1548 (2012)
5. Cozzolino, A., Flammini, F., Galli, V., Lamberti, M., Poggi, G., Pragliola, C.: Evaluating the effects of MJPEG compression on Motion Tracking in metro railway surveillance. In: Blanc-Talon, J., Philips, W., Popescu, D., Scheunders, P., Zemčık, P. (eds.) ACIVS 2012. LNCS, vol. 7517, pp. 142–154. Springer, Heidelberg (2012)
6. Liu, X., Tu, P.H., Rittscher, J., Perera, A., Krahnstoeber, N.: Detecting and counting people in surveillance applications. In: Advanced Video and Signal Based Surveillance, AVSS (2005)
7. Zhang, L., Wu, B., Nevatia, R.: Pedestrian detection in infrared images based on local shape features. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2007, pp. 1–8. IEEE (2007)
8. <http://www.axis.com/>
9. Regazzoni, C.S., Cavallaro, A., Wu, Y., Konrad, J., Hampapur, A.: Video analytics for surveillance: Theory and practice. *IEEE Signal Processing Magazine* 27(5), 16–17 (2010)
10. Bocchetti, G., Flammini, F., Pragliola, C., Pappalardo, A.: Dependable integrated surveillance systems for the physical security of metro railways. In: Third ACM/IEEE International Conference on Distributed Smart Cameras, ICDS 2009, pp. 1–7 (2009)
11. Rahmalan, H., Nixon, M.S., Carter, J.N.: On crowd density estimation for surveillance. In: Crime and Security (2006)
12. Duque, D., Santos, H., Cortez, P.: Prediction of abnormal behaviors for intelligent video surveillance systems. In: Computational Intelligence and Data Mining (2007)