Available Seat Counting in Public Rail Transport

Pieterjan De Potter, Christophe Billiet, Chris Poppe, Brecht Stubbe, Steven Verstockt, Peter Lambert, Rik Van de Walle

Proceedings of the 29th Progress in Electromagnetics Research Symposium (PIERS), pp. 1294-1298, 2011.

To refer to or to cite this work, please use the citation to the published version:

Pieterjan De Potter, Christophe Billiet, Chris Poppe, Brecht Stubbe, Steven Verstockt, Peter Lambert, Rik Van de Walle (2011). Available Seat Counting in Public Rail Transport. *Proceedings of the 29th Progress in Electromagnetics Research Symposium (PIERS)*, pp. 1294-1298.

Available Seat Counting in Public Rail Transport

P. De Potter[1], **C. Billiet**[1], **C. Poppe**[1], **B. Stubbe**[2], **S. Verstockt**[1,3], **P. Lambert**[1], **and R. Van de Walle**[1]

[1]Multimedia Lab, Department of Electronics and Information Systems, Ghent University-IBBT, Belgium
[2]Televic Group NV, Belgium
[3]ELIT Lab, University College West Flanders, Ghent University Association, Belgium

Abstract— Surveillance cameras are found almost everywhere today, including vehicles for public transport. A lot of research has already been done on video analysis in open spaces. However, the conditions in a vehicle for public transport differ from these in open spaces, as described in detail in this paper. A use case described in this paper is on counting the available seats in a vehicle using surveillance cameras. We propose an algorithm based on Laplace edge detection, combined with background subtraction.

1. INTRODUCTION

Video analysis has become a necessity because of the never ending growth of the amount of video surveillance cameras. Different intelligent video surveillance systems have already been developed for the detection of unwanted events [1–3]. In public transportation as well, video surveillance cameras have made their entrance, and thus video analysis is needed. However, because the conditions in vehicles for public transportation are different, a different approach is necessary.

The primary goal of video surveillance cameras is of course video surveillance, but if the cameras are installed, they can also be used for other video analysis tasks. One of these tasks is available seat counting, which makes it possible to refer passengers to cars with empty seats and to gather statistical information.

In this paper, we propose a system that combines an illumination change invariant edge detection technique with a background subtraction technique that can detect the blobs representing the moving objects.

The remainder of this paper is organized as follows. In Section 2, we discuss some of the related work in this area. We elaborate on the specific conditions in vehicles for public transportation in Section 3. In Section 4, an approach is presented to count the available seats in a vehicle. An evaluation of this implementation is given in Section 5. Finally, conclusions and future work are given in Section 6.

2. RELATED WORK

While a lot of recent research is done on the topic of video analysis, the number of publications in the area of analysis inside moving vehicles is quite limited.

In [4], Milcent et al. present a system to detect baggage in transit vehicles. They preprocess the video stream to correct the lighting. A light location mask, indicating reflecting metallic posts inside the vehicle, is used to gather the different parts of one object. To increase the speed of the segmentation algorithm, it is only applied on a region indicated by a probability location mask.

Several projects, such as PRISMATICA (pro-active integrated systems for security management by technological, institutional and communication assistance [5]) and BOSS (on board wireless secured video surveillance [6]) mention the transmission of video feeds upon the triggering of an alarm, but do not describe how the alarm is exactly triggered.

In [7], Vu et al. present an event recognition system based on face detection and tracking combined with audio analysis. 3D context such as zones of interest and static objects are recorded in a knowledge base and 3D positions are calculated for mobile objects using calibration matrices. Strong changes in lighting conditions occasionally prevent the system to detect people correctly.

Yayahiaoui et al. [8] and Liu et al. [9] report a high accuracy in passenger counting using a dedicated setup. Since the cameras used for this setup can not be used for other purposes, this solution is too expensive to be used in real life. Also, it is impossible to retrieve the location of the passengers.

3. CONDITIONS IN VEHICLES FOR PUBLIC TRANSPORTATION

In this section, we elaborate on the conditions that are specific for vehicles for public transportation.

One of the big challenges in video analysis is dealing with illumination variances and shadows. Inside moving and turning vehicles, this problem becomes even worse. The angle of the incoming sunlight changes, driving into a tunnel reduces the illumination drastically and static and moving objects almost always cast a dark shadow.

Another problem that is enlarged in vehicles is that of occlusion. Because of the limited space, moving objects easily occlude each other and the biggest part of seated passengers cannot be seen because of occlusion by the seats.

The windows pose other problems: during the day, a fast moving background can be observed through this windows. When it gets darker outside, objects are reflected.

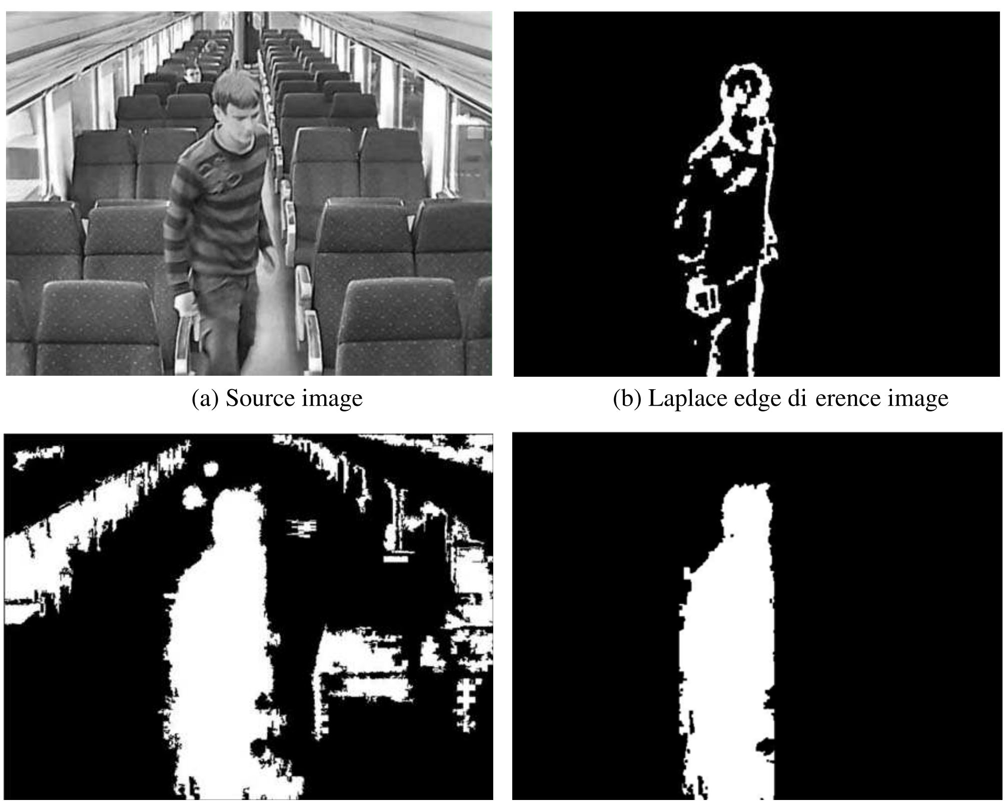
Since all the equipment has to be installed on the train, it has some limitations: the bandwidth between different pieces of equipment is low, the available processing power is limited and the installed cameras provide low quality video streams containing a lot of noise.

4. AN APPROACH FOR AVAILABLE SEAT COUNTING

In this section, our approach to count the number of occupied seats, from which the number of available seats can be derived, is described. In a first subsection, we describe how we detect moving objects using different methods. In a second subsection, we introduce the event detection that leads to the number of occupied seats.

4.1. Object Detection

The classification of pixels in foreground and background pixels is done in the object detection phase. We use different techniques for the object detection, as it has already been proven in the past that the combination of multiple techniques can reduce the individual weaknesses of these techniques [10]. The object detection consists of three consecutive steps: first, an edge detector is applied to discover the contours of moving objects. Secondly, a background subtraction method is used to retrieve blobs of potential foreground objects. A last step consists of merging the results



(a) Source image

(b) Laplace edge di erence image

(c) Background subtraction

(d) Combined result

Figure 1: Object detection.

of both techniques to obtain the blobs of the actual foreground objects. In this subsection, these three steps are described in more detail and illustrated using the image depicted in Figure 1(a).

To detect the contours of moving objects, a Laplace edge detector with a 7×7 pixel mask is used to extract the edges in a frame. The size of the mask is not too big, in order to keep the computation time within limits. The resulting edge image is then compared to the edge image of the previous frame to obtain a difference image containing the contours of moving objects. An example of such a difference edge image is given in Figure 1(b). The edge detector is quite independent on changes in illumination, since it only considers 2 frames.

For the background subtraction method, in a training phase the median value is calculated for every RGB channel of each pixel, as well as the average deviation from this median value. During the object detection phase, a pixel is then assumed to be a foreground pixel if its value differs more than 11 times the average deviation from the median value in each RGB channel. In a post processing phase, noise is eliminated by removing small connected areas. The areas that remain are interpreted as foreground areas and holes in these areas are filled up. The result of the background subtraction method is illustrated in Figure 1(c).

The results of both techniques are merged by taking the result of the background subtraction method and removing all the blobs that do not contain a significant amount of edges in the result of the edge detector method. By doing this, only the blobs corresponding to moving objects remain, as shown in Figure 1(d). It can be seen that only the moving, standing person is detected as a moving object, while the sitting persons in the upper, left-hand side of the image, are not detected as moving objects. This is not necessary, since they stay seated and thus no events need to be detected by the event detection mechanism.

4.2. Event Detection

The counting algorithm is based on the following principle: the number of seated passengers can only be increased or decreased when a passenger sits down or leaves a seat. The total number of seated passengers can thus be determined by counting the sit and leave actions.

In order to count these actions, multiple rectangular regions are identified on a camera view of the vehicle, as illustrated in Figure 2(a). These rectangular regions can be divided into three groups: the regions representing the left-hand side seats, the aisle and the right-hand side seats. The regions in the seats groups are split up in smaller rectangles, named tiles below, as illustrated in Figure 2(b), to improve the results.

The creation of the rectangles and tiles is done manually, but since it is a static environment and only a couple of seat configurations exist, this only has to be done a few times.

A tile is triggered when at least half of its pixels are detected as foreground pixels. When half of the tiles of a rectangle are triggered, the rectangle is triggered and sit action detection is started. The order in which the tiles were triggered is checked and is compared with previous presence of foreground pixels in either the aisle or an adjacent seat region, depending on the situation. If the combination of the order in which the tiles are triggered and previous presence in adjacent regions makes sense, then a sit activity is registered. E.g., If the tiles in the region from a seat near the aisle are triggered in an order from the the aisle to the window and previous presence was registered in

(a) Rectangular regions: seats and aisle

(b) Further splitting up of the seat regions in tiles

Figure 2: Identification of different regions in a camera view.

(a) Group 1: One actor (b) Group 2: Multiple actors not acting in group (c) Group 3: Multiple actors acting in group

Figure 3: The different categories of test sequences.

Table 1: Performance for the different groups of video sequences.

Group	Precision	Recall
1	100%	100%
2	100%	57.14%
3	85.71%	85.71%

the aisle, then a sit action is registered for one of the seats near the aisle. If the comparison makes no sense, then no action is taken.

A rectangle can also become untriggered after being triggered before, when at least half of the tiles are untriggered, that is when at most half of its pixels are detected as foreground pixels, after being triggered before. Again, the order in which the tiles are untriggered is checked and compared to presence in adjacent regions. In this situation, a leave action is registered when the combination of the order and presence makes sense.

5. EVALUATION

The performance of our algorithm was tested on 78 acted sequences recorded in a passenger coach provided by the NMBS (national railway company of Belgium). The resolution of the sequences is 640×480 pixels and the framerate is 25 fps. The sequences are divided by difficulty into three groups: in the first group, only one actor is present. In the second group, multiple actors are present, but they do not perform group actions. In the third group, multiple actors are present and they do perform actions, such as entering the vehicle and sitting down, in group. An example of each of these categories can be seen in Figure 3.

The ground truth for these sequences is build up using the knowledge of the actual number of persons sitting down at a certain moment and can thus differ from the number of persons sitting down as observed at a certain moment because of occlusion.

The accuracy of the proposed algorithm is given in Table 1, by the values for precision and recall for the different groups of sequences. Unfortunately, currently no datasets and results are available to compare with.

The lower recall value for group 2 then for group 3 can be explained by the fact that, when people want to go sit together, they have to wait turns before moving into their seats. When people do not move in group, there is a bigger chance they perform a sit or leave action at the same time, which can lead to an occlusion.

The algorithm has not yet been optimized for speed, but manages to process one frame every 60 milliseconds on a 2.2 GHz processor. Some speed optimizations are thus required to meet real time constrains, for which a maximum processing time of 40 milliseconds is demanded for a single frame of the 25 fps sequences. Of course, more adaptations would be needed to be able to run this software on embedded hardware.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we described the conditions that are specific for vehicles in public rail transport and give an approach to count the number of seated passengers under these conditions. The proposed solution performs quite well in terms of precision and recall, but would need adaptations before being able to run on embedded hardware.

One thought that came into mind was that it might be useful not only to use the same hardware as the video surveillance system, but also a part of the software. For example the results of the object detection algorithm could be shared. This could lead to a minimal extra load for additional video analysis tasks.

Future work consists of trying other methods for a robuster object detection, so that also other events, such as fights, could be detected. Also, other camera configurations, with cameras in the middle of the aisle can be investigated. The dataset we used for the evaluation will be published to make a comparison of our results with other results possible.

ACKNOWLEDGMENT

The research activities as described in this paper were funded by Ghent University, the Interdisciplinary Institute for Broadband Technology (IBBT), the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT), the Fund for Scientific Research-Flanders (FWO-Flanders), and the European Union.

REFERENCES

1. Haritaoglu, I., D. Harwood, and L. S. Davis, "W4: Real-time surveillance of people and their activities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, 809–830, 2000.
2. Bremond, F., M. Thonnat, and M. Zuniga, "Video understanding framework for automatic behavior recognition," *Behavior Research Methods*, Vol. 36, No. 3, 416–426, 2006.
3. Adam, A., E. Rivlinand, I. Shimshoni, and D. Reinitz, "Robust real-time unusual event detection using multiple fixed-location monitors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 30, No. 3, 555–560, 2008.
4. Milcent, G. and Y. Cai, "Location based baggage detection for transit vehicles," 15 p, Carnegie Mellon University, CyLab, Pittsburgh (PA), December 2005.
5. Velastin, S. A., L. Khoudour, B. P. L. Lo, J. Sun, and M. A. Vicencio-Silva, "PRISMATICA: A multi-sensor surveillance system for public transport networks," *Proceedings of 12th IEE International Conference on Road Transport Information and Control*, 19–25, London, UK, April 2004.
6. Jeney, G., C. Lamy-Bergot, X. Desurmont, R. Lopez da Silva, R. A. García-Sanchidrián, M. Bonte, M. Berbineau, M. Csapodi, O. Cantineau, N. Malouch, D. Sanz, and J.-L. Bruyelle, "Communications challenges in the celtic-BOSS project," *Lecture Notes in Computer Science*, Vol. 4712, 431–442, 2007.
7. Vu, V.-T., Q.-C. Pham, J.-l. Rouas, F. Bremond, N. Allezard, S. Ambellouis, G. Davini, P. Sayd, A. Flancquart, and M. Thonnat, "Audio-video event recognition system for public transport security," *Proceedings of IET Conference on Imaging for Crime Detection and Prevention*, 414–419, London, UK, June 2006.
8. Yahiaoui, T., C. Meurie, L. Khoudour, and F. Cabestaing, "A people counting system based on dense and close stereovision," *Lecture Notes in Computer Science*, Vol. 5099, 59–66, 2008.
9. Liu, N. and C. Gao, "Bi-directional passenger counting on crowded situation based on sequence color images," *Lecture Notes in Computer Science*, Vol. 4282, 557–564, 2006.
10. Poppe, C., G. Martens, S. De Bruyne, P. Lambert, and R. Van de Walle, "Robust spatio-temporal multimodal background subtraction for video surveillance," *Optical Engineering*, Vol. 47, No. 10, 107203, 2008.