

biblio.ugent.be

The UGent Institutional Repository is the electronic archiving and dissemination platform for all UGent research publications. Ghent University has implemented a mandate stipulating that all academic publications of UGent researchers should be deposited and archived in this repository. Except for items where current copyright restrictions apply, these papers are available in Open Access.

This item is the archived peer-reviewed author-version of:

Automatic Available Seat Counting In Public Rail Transport Using Wavelets

Pieterjan De Potter, Ioannis Kypraios, Steven Verstockt, Chris Poppe, Rik Van de Walle

Proceedings of the 53rd International Symposium ELMAR

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

Pieterjan De Potter, Ioannis Kypraios, Steven Verstockt, Chris Poppe, Rik Van de Walle (2011). Automatic Available Seat Counting In Public Rail Transport Using Wavelets. *Proceedings of the 53rd International Symposium ELMAR*, pp. 79-83.

Automatic Available Seat Counting In Public Rail Transport Using Wavelets

Pieterjan De Potter¹, Ioannis Kypraios², Steven Verstockt^{1,3}, Chris Poppe¹, Rik Van de Walle¹

¹Department of Electronics and Information Systems, Multimedia Lab, Ghent University – IBBT

Gaston Crommenlaan 8, bus 201, B-9050 Ledeberg-Ghent, Belgium

²School of Engineering and Design, University of Sussex

Falmer Brighton, BN1 9QT, United Kingdom

³ELIT Lab, University College West Flanders, Ghent University Association

Graaf Karel de Goedelaan 5, 8500 Kortrijk, Belgium

E-mail: Pieterjan.DePotter@UGent.be

Abstract—Previously, we introduced an available seat counting algorithm in public rail transport. The main disadvantage of that algorithm is that it lacks automatic event detection. In this paper, we implement two automatic wavelet-based available seat counting algorithms. The new algorithms employ the spatial-domain Laplacian-of-Gaussian-based wavelet, and the frequency-domain Non-Linear Difference of Gaussians-based wavelet bandpass video scene filter to extract illumination invariant scene features and to combine them efficiently into the background reference frame. Manual segmentation of the scene into rectangles and tiles to detect seated objects is no longer needed as we now apply a boundary box tracker on the segmented moving objects' blobs. We test all the algorithms with different video sequences in passengers' train coaches, and compare the previous approach with the two new automatic wavelet-based available seat counting algorithms, and an additional spatial-domain automatic non-wavelet based Simple Mixture of Gaussian Models.

Keywords—video analysis, seat counting, public transport, wavelets

I. INTRODUCTION

Over the past decade, the number of installed video surveillance cameras has grown exponentially because of the reduced cost and the fact that security has gained importance over privacy in some scenarios. This has led to the development of different video analytics systems to detect different scenarios' events [1], [2], [3]. In public transport as well, video surveillance cameras are being installed, and video analytics are becoming helpful. However, the different conditions in vehicles turn the video analytics' task difficult. The primary goal is to provide additional security, but as the cameras are already installed, they can also be used for other purposes such as seat counting.

While a lot of research has already been conducted on the topic of video analytics, the number of publications for scenarios 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. Three dimensional (3-D) context such as zones of interest and static objects are recorded in a knowledge base and 3-D positions are calculated for mobile objects using calibration matrices. Strong changes in lighting conditions occasionally prevent the system to detect people correctly. Yahiaoui et al. [8] and Liu et al. [9] report high accuracies 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 some real life scenarios. Also, it is impossible to retrieve the location of the passengers.

In a previous paper [10], we proposed a system to tackle the problem of seat counting. The main disadvantages are that manual labor is needed for each camera view and a training phase is necessary. In this paper, we propose two automatic wavelet-based available seat counting algorithms that extract and combine illumination invariant scene features efficiently into their composed background reference frame.

The remainder of this paper is organized as follows. In Section II, we discuss our previous work of a non-automatic available seat counting algorithm. In Section III, we describe the two wavelet based available seat counting algorithms. An evaluation of the previously described non-automatic algorithm, the two algorithms described in this paper and a Simple Mixture of Gaussian Models (SMM) [11] based algorithm is given in Section IV. Finally, conclusions and future work are given in Section V.

II. NON-AUTOMATIC AVAILABLE SEAT COUNTING

In [10], we presented an approach to tackle the available seat counting problem. This approach consists of two stages: object detection and event detection.

The object detection consists of three consecutive steps: first, Laplacian edge detection is applied to discover the



(a) Camera 1 (b) Camera 2

Fig. 1. Sample image from the test sequences

contours of moving objects. Secondly, a median based background subtraction method is used to retrieve blobs of potential foreground objects. A last step consists of merging the results of both techniques to obtain the blobs of the actual foreground objects.

In the event detection stage, *sit down* and *leave* actions are detected to obtain the number of available seats in a vehicle. For this purpose, rectangular regions are defined manually at the positions of the seats. This rectangles are further subdivided in manually defined tiles. 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 compared with previous presence of foreground pixels in either the aisle or an adjacent seat region. *Sit down* activity is registered when aisle or adjacent seat foreground pixels are detected triggering the seat tiles. For leave seat action detection, an opposite process is executed. Another drawback of this algorithm is that Camera 1 (CAM1) (see Fig. 1) can process only half of the passengers' coach and Camera 2 (CAM2) can process the other half of the passengers' coach. This is due to the limitations of the manually defined rectangles and tiles with respect to the perspective ratio of the passengers' coach.

III. WAVELET-BASED AUTOMATIC AVAILABLE SEAT COUNTING

A. Laplacian of Gaussian

In our previously described algorithm, we combined edge detection with a background subtraction method. Now, since we want more robustness against illumination changes, and the previously applied background subtraction method is too computationally expensive and needs a training phase, here we only apply edge detection.

From (1), on each frame we first apply a Gaussian filter $G(x, y)$ in the spatial domain to cope with the noise in the image. The variance σ of the filter is chosen to be the same in x- and y- direction and dependent on the kernel size. Then in (2), we apply a Laplacian filter $L(x, y)$, again in the spatial domain, to detect the edges. This results in the Laplacian-of-Gaussian (LoG) operation $LoG(x, y)$, which is shown in (3). It can be shown that the LoG operation acts as a bandpass filter. By selecting the right kernel dimensions, which for our test video sequences was found to be a 7x7 size kernel, the

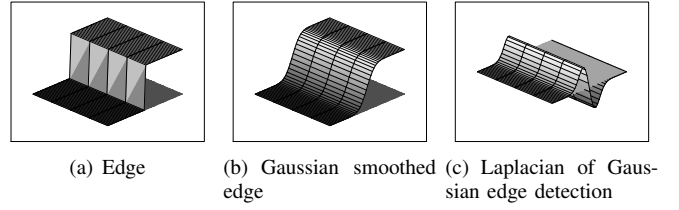


Fig. 2. Laplacian of Gaussian operator applied to an edge; the zero-plane is also plotted

background noise can be filtered out almost completely, while maintaining the edges.

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (1)$$

$$L(x, y) = \frac{\partial^2 f(x, y)}{\partial x^2} + \frac{\partial^2 f(x, y)}{\partial y^2} \quad (2)$$

$$LoG(x, y) = \frac{x^2 + y^2 - 2\sigma^2}{2\pi\sigma^6} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (3)$$

When the LoG operation is applied to the image edges, it produces positive values at the one side of the edge and negative values at the other side (see Fig. 2). Hence we check the result of the LoG operation for zero-crossings in the horizontal, vertical and both diagonal directions to obtain the edges. The value 0 is returned if the adjacent edge values have the same sign; the absolute of the difference of the edge values is returned when they have an opposite sign. For each pixel, the maximum edge value over all four directions is given as the final result. The results are for each frame subtracted from the results from the previous frame to obtain the moving edges in the current frame.

B. Non-Linear Difference of Gaussians

By subtracting two Gaussian concentric kernels with different standard deviation values for a limited time duration, a new kernel is formed [12]. This kernel has an average value of zero and is useful for wavelet analysis applications. The resulted difference of Gaussians (DoG) filter can detect edges independent of orientation and produce, when applied, an edged enhanced image [12]. This operation is given by (4), where $g_1(x, y, \sigma_1)$ and $g_2(x, y, \sigma_2)$ are the two Gaussian kernels with standard deviations σ_1 and σ_2 , $\Phi(x, y)$ is the input image (in spatial domain), and $\Phi^{DoG}(x, y)$ is the linearly (linear difference) convolution image with the two Gaussian kernels.

$$\Phi^{DoG}(x, y) = (\Phi(x, y) \otimes g_1(x, y, \sigma_1)) - (\Phi(x, y) \otimes g_2(x, y, \sigma_2)) \quad (4)$$

It can be found that the DoG filter forms a type of band-pass filter with lower and upper cutoff frequencies set by the two Gaussian kernels. By tuning the standard deviation parameters σ_1 and σ_2 , the DoG filter is able to select the discriminative passband mid-frequency features for scene objects and to stop

the low-frequency illumination changes effects and any high-frequency noise in the input image scene [13]. It is proven that the DoG filter approximates the Laplacian ∇^2 operator (or the two-dimensional second directional derivative of the Gaussian kernels ∇^2 that is used to create a narrow band-pass differential operator [13]) best when the ratio of the two standard deviations σ_1/σ_2 is equal to 1.6.

The following observations can be made on the DoG filter operation: (a) the DoG ∇^2 operator creates a non-uniform distribution of energy around the image it is applied on [14]; (b) the partially closed areas of the image have higher energy levels relative to other areas. Thus, this unequal energy distribution causes the image being highly sensitive to rotation and scale changes of edges. In [14], the authors have shown that by applying a non-linear function on top of the DoG ∇^2 operator, the produced non-linear DoG (NL-DoG) filter allows a more uniform distribution of energy around the closed regions of the image. In practice, this causes more fine details of the image around the edges to be enhanced. The non-linear function \aleph is applied in the spatial domain of the image [14]. When \aleph is applied on top of the DoG ∇^2 operator, the resulting image $\Phi^{NL-DoG}(x, y)$ is given by (5).

$$\Phi^{NL-DoG}(x, y) = \aleph \cdot \Phi^{DoG}(x, y) \quad (5)$$

For the results shown in this paper, we have applied the NL-DoG filter in the frequency domain. Hence, (4) becomes (6), where the Fast Fourier Transform operation is shown as FFT and $\Phi_{FFT}^{DoG}(x, y)$ is the DoG filter transformed image $\Phi^{DoG}(x, y)$ in the frequency domain. Then, (5) can be rewritten as (7), where the inverse-fast Fourier transform is shown as $IFFT$. Thus, \aleph is applied in the spatial domain but the rest of the NL-DoG filter is applied in the frequency domain. \aleph is chosen to be a sigmoidal-type function.

$$\Phi_{FFT}^{DoG}(x, y) = FFT(\Phi(x, y)) \cdot FFT(g_1(x, y, \sigma_1) - g_2(x, y, \sigma_2)) \quad (6)$$

$$\Phi^{NL-DoG}(x, y) = \aleph \cdot IFFT(\Phi_{FFT}^{DoG}(x, y)) \quad (7)$$

Fig. 3 shows the NL-DoG filter implementation for scene segmentation of foreground objects in the automatic available seat counting algorithm we have developed. We used the time intervals with memory (TIME, [15]) algorithm to compose the reference background frame. Thus, a background frame is selected at regular time intervals for the whole duration of each test video sequence recorded from CAM1 or CAM2 installed in the passengers' train coaches. In Fig. 3, assume the test video sequence we used has a total duration of 35 seconds with 13 frames per second (fps) (after clearing out the duplicate frame patterns produced during the video data acquisition and storing), then we select one frame per fixed time interval each second i.e. in total 35 frames were selected for the composition of the reference background frame. Then, NL-DoG filter is applied on each selected frame in the frequency domain, and all the NL-DoG transformed frames are averaged to synthesize

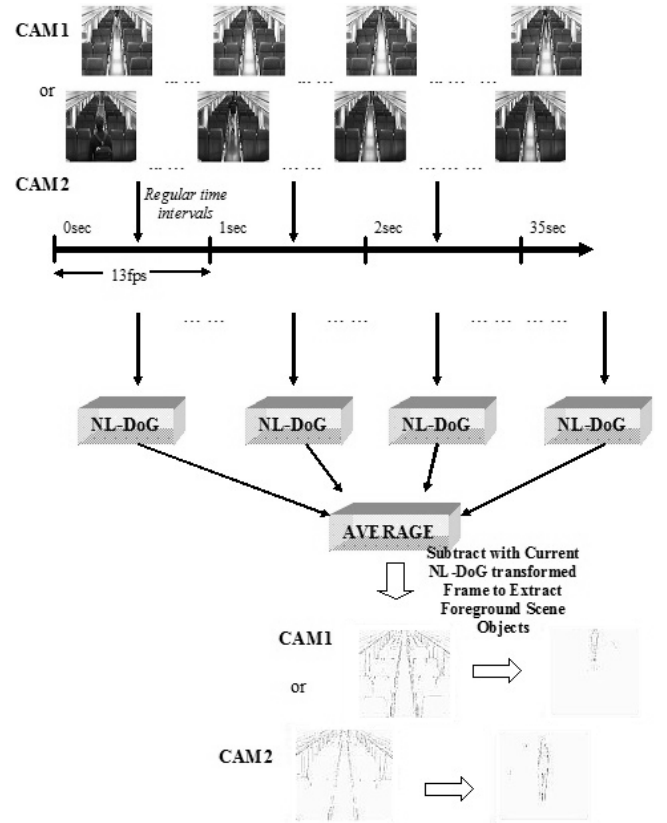


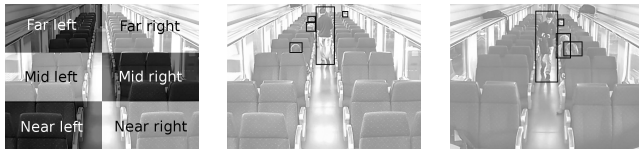
Fig. 3. Foreground scene objects extraction in the test video sequences recorded from Camera 1 and 2, installed in the passengers' train coaches, using the TIME algorithm and the NL-DoG wavelet-based filter

the background reference frame. We apply the NL-DoG filter in the frequency domain on the current test video sequence frame and subtract the composed background reference frame to extract the foreground scene objects [16].

C. Bounding Box Tracking

To achieve automatic available seat counting in our developed algorithms, we implemented a simple Bounding Box (BB) tracking method in combination with the LoG (spatial domain), NL-DoG (frequency domain), and the simple mixture of gaussians (SMM, [11]) (spatial domain) operations.

In a first step, BBs are applied on the detected objects. A threshold value is used for the LoG/NL-DoG/SMM transformed images (or video frames) of the algorithms, followed by a morphological closure and opening. The produced objects' blobs are compared with those in the previous frame. If there is a similar object blob in the previous frame, BB is matched and passed on to the next step. In a second step, BBs are eliminated based on multiple criteria concerning their size relative to their position and the perspective ratio in the passengers' coach. Thus, detected objects further away from the camera are assumed to be smaller than objects closer to the camera. The remaining BBs are used in the final step.



(a) Passenger coach set-tings independent scene map (b) Bounding boxes results for camera 1 (c) Bounding boxes results for camera 2

Fig. 4. Scene map and bounding box results

Here, we implemented a novel scene map [17] independent of the different passengers' train coaches. In effect, we apply a scene map on the passengers' coach, where we assume the coach can be divided to FAR-LEFT, FAR-RIGHT, MID-LEFT, MID-RIGHT, NEAR-LEFT, NEAR-RIGHT (see Fig. 4(a)), or simplified to LEFT and RIGHT zones. No explicit pixel counting is needed to separate into different zones. Rather, the created scene map is independent on any settings of the passengers' train coaches. Therefore, in the final step, detected objects (see Fig. 4(b) and Fig. 4(c)) are classified to be occupying a seat when they are moving into the LEFT or RIGHT zones of the map.

IV. EVALUATION

A. Test Setup

We evaluated the different algorithms on sequences that were recorded in a train of the Belgian national railway company (NMBS-Group). Two cameras, CAM1 and CAM2, installed in the passengers' train coach were used to record these video sequences; sample video frames are shown in Fig. 1(a) for CAM1 and Fig. 1(a) for CAM2. We preprocessed the recorded video sequences to clear out any duplicate patterns of frames created during the acquisition and storing stage of the sequences.

B. Evaluation Metrics

For each frame in the sequences, the actual number of persons in the seats of the train is compared with the number given by the different approaches. The minimum of these two numbers are counted as true positives (TP); if the actual number of persons is greater than the detected number, this excess is counted as false negatives (FN); if the actual number of persons is less, this shortage is counted as false positives (FP). The true negatives (TN) consist of the number of frames that are correctly detected as frames with no persons present.

Based on these metrics, the precision, recall, true negative rate, and accuracy are calculated as follows:

$$\text{precision} = \frac{TP}{TP + FP} \quad (8)$$

$$\text{recall} = \frac{TP}{TP + FN} \quad (9)$$

$$\text{true negative rate} = \frac{TN}{TN + FP} \quad (10)$$

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

C. Preliminary Results

Table I summarizes the recorded performance metrics for the conducted initial results of the different algorithms, i.e. the non-automatic available seat counting algorithm, and the new automatic wavelet-based LoG/NL-DoG and non-wavelet based SMM available seat counting algorithms. Notice we have recorded the performance metrics for both CAM1 and CAM2. For our tests, we used two video sequences, one of 35 seconds and one of approximately 4 minutes. For the non-automatic available seat counting algorithm, we only show one value, instead of two (one for each camera). This is due to the limitations of this algorithm in which each camera can only process half of the passengers' coach. So, the value shown is the combined result of the two cameras.

From the recorded precision and accuracy performance metrics values, it is shown that the automatic wavelet-based LoG/NL-DoG and non-wavelet based SMM algorithms perform superior to the non-automatic available seat counting algorithm. In effect, the automatic available seat counting algorithms are able to produce a higher number of TPs than the non-automatic available seat counting algorithm. From the recorded recall values, it can be found that the non-automatic available seat counting algorithm exhibits a higher value than the automatic (wavelet- or non-wavelet based) available seat counting algorithms. In effect, focusing on the long test video sequences, and if we take the average value of CAM1 and CAM2, then we get approximately 0.81 for LoG, 0.78 for NL-DoG and 0.62 for SMM. This means that the non-automatic available seat counting algorithm has less time during which a seated person is not detected than the automatic available seat counting algorithms. On the other hand, if we take a look at the true negative rate, we obtain an average value of approximately 0.25 for LoG, 0.17 for NL-DoG and 0.71 for SMM, but 0.03 for the non-automatic algorithm. This means that the non-automatic available seat counting algorithm has more time during which a seated person is falsely detected.

V. CONCLUSION AND FUTURE WORK

This paper described two new automatic wavelet-based available seat counting algorithms. One was based on spatial domain LoG, and the other on the frequency domain NL-DoG. They both are able to select illumination invariant features in the synthesis of their background reference frame. A novel scene map method together with a BB tracking method is used to detect and classify the segmented objects as seated. From the recorded initial results, it is shown that the automatic available seat counting algorithms in spatial- and frequency-domains outperform the non-automatic available seat counting algorithm.

In future, we are going to experiment with the creation of a robust tracking mask to recognize multiple objects and to separate them into different categories [18], [19]. Also, a major challenge to solve remains dealing with objects' occlusions in

TABLE I
PERFORMANCE EVALUATION OF DIFFERENT APPROACHES FOR AVAILABLE
SEAT COUNTING

Algorithm	Precision	Recall	True negative rate	Accuracy
Previous approach				
<i>Short</i>	1	0.8325	1	0.8839
<i>Long</i>	0.6366	0.9940	0.0274	0.6379
Laplacian of Gaussian				
<i>Short cam 1</i>	0.9918	0.9032	0.9816	0.9258
<i>Short cam 2</i>	1	0.8846	1	0.9162
<i>Long cam 1</i>	0.9260	0.8468	0.1653	0.7957
<i>Long cam 2</i>	0.9706	0.7819	0.3429	0.7667
non-linear difference of Gaussians				
<i>Short cam 1</i>	1	0.5146	1	0.603
<i>Short cam 2</i>	1	0.5529	1	0.9679
<i>Long cam 1</i>	0.9533	0.6641	0.2446	0.6467
<i>Long cam 2</i>	0.8570	0.8867	0.0866	0.7752
Simple mixture of models				
<i>Short cam 1</i>	1	0.5326	1	0.6204
<i>Short cam 2</i>	1	0.5142	1	0.6015
<i>Long cam 1</i>	0.9944	0.6002	0.7373	0.6020
<i>Long cam 2</i>	0.9926	0.6301	0.6788	0.6311

the scene. We will create more tests video datasets to further compare the performance of LoG-based, NL-DoG based, and SMM-based automatic available seat counting algorithms combined with the tracking mask.

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] I. Haritaoglu, D. Harwood, and L. Davis, "W-4: Real-time surveillance of people and their activities," *IEEE Trans. on Pattern Anal. and Mach. Intell.*, vol. 22, no. 8, pp. 809–830, Aug. 2000.
[2] F. Bremond, M. Thonnat, and M. Zuniga, "Video-understanding framework for automatic behavior recognition," *Behav. Res. Methods*, vol. 38, no. 3, pp. 416–426, Aug. 2006.

[3] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz, "Robust real-time unusual event detection using multiple fixed-location monitors," *IEEE Trans. on Pattern Anal. and Mach. Intell.*, vol. 30, no. 3, pp. 555–560, Mar. 2008.
[4] G. Milcent and Y. Cai, "Location Based Baggage Detection for Transit Vehicles," *Carnegie Mellon University, Tech. Rep.*, Dec. 2005.
[5] S. Velastin, L. Khoudour, B. Lo, J. Sun, and M. Vicencio-Silva, "PRISMATICA: a multi-sensor surveillance system for public transport networks," in *12th IEE Int. Conf. on Road Transp. Inf. & Control - RTIC*. IEE, Apr. 2004, pp. 19–25.
[6] G. Jeney, C. Lamy-Bergot, X. Desurmont, R. da Silva, R. Garcfa-Sanchidrian, M. Bonte, M. Berbineau, M. Csapodi, O. Cantineau, N. Malouch, D. Sanz, and J.-L. Bruyelle, "Communications challenges in the Celtic-BOSS project," in *Next Gener. Teletraffic and Wired/Wirel. Adv. Netw., Proc. 7th Int. Conf., NEW2AN 2007.. (Lect. Notes in Comp. Sci. vol. 4712)*, Sep. 2007, pp. 431–42.
[7] V.-T. Vu, F. Bremond, G. Davini, M. Thonnat, Q.-C. Pham, N. Allezard, P. Sayd, J.-L. Rouas, S. Ambellouis, and A. Flancquart, "Audio-video event recognition system for public transport security," in *IET Conf. on Crime and Secur.*, Jun. 2006, p. 6 pp.
[8] T. Yahiaoui, C. Meurie, L. Khoudour, and F. Cabestaing, "A people counting system based on dense and close stereovision," in *Image and Signal Process. 3rd Int. Conf., ICISP*, Jul. 2008, pp. 59–66.
[9] N. Liu and C. Gao, "Bi-directional passenger counting on crowded situation based on sequence color images," in *Adv. in Artif. Reality and Tele-Exist. 16th Int. Conf. Proc. (Lect. Notes in Comp. Sci. Vol. 4282)*, Nov. 2006, pp. 557–64.
[10] P. De Potter, C. Billiet, C. Poppe, B. Stubbe, S. Verstockt, P. Lambert, and R. Van de Walle, "Available Seat Counting in Public Rail Transport," in *Proc. PIERS 2011, Marakkech*, Mar. 2011, pp. 1294–1298.
[11] C. Poppe, G. Martens, S. De Bruyne, P. Lambert, and R. Van de Walle, "Robust spatio-temporal multimodal background subtraction for video surveillance," *Opt. Eng.*, vol. 47, no. 10, Oct. 2008.
[12] D. Marr and E. Hildreth, "Theory of Edge-detection," *Proc. R. Soc. of Lond. Ser. B-Biol. Sci.*, vol. 207, no. 1167, pp. 187–217, Feb. 1980.
[13] O. Arandjelovic and R. Cipolla, "Achieving Illumination Invariance using Image Filters," in *Face Recognition*, K. Delac and M. Grgic, Eds. I-Tech Education and Publishing, Jul. 2007, pp. 15–30.
[14] L. Jamal-Aldin, R. Young, and C. Chatwin, "Application of nonlinearity to wavelet-transformed images to improve correlation filter performance," *Appl. Opt.*, vol. 36, no. 35, pp. 9212–9224, Dec. 1997.
[15] M. H. Khan, I. Kypraios, and U. Khan, "A robust background subtraction algorithm for motion based video scene segmentation in embedded platforms," in *Front. in Inf. Technol. FIT-2009, ACM, Pakistan*, Dec. 2009.
[16] S. Verstockt, I. Kypraios, P. De Potter, C. Poppe, and R. Van de Walle, "Wavelet-based Multi-modal Fire Detection," unpublished.
[17] I. Kypraios, "A multi-level alarm algorithm for recognising human movement and detecting abandoned baggages," unpublished.
[18] I. Kypraios, "Video analytics algorithms for smart ip camera for i-lids sterile zone scenario solution of human synthetic movement monitoring," 2020Imaging ltd. U.K., Tech. Rep., Dec. 2008, algorithms for an Embed. Microprocess., TR/2020/02081202.
[19] I. Kypraios, "Monitoring synthetic human movement for sterile zone scenario," unpublished.