Online ISSN 1848-3380, Print ISSN 0005-1144
ATKAFF 53(4), 321–334(2012)

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

Automatic Passengers Counting In Public Rail Transport Using Wavelets

DOI 10.7305/automatika.53-4.227
UDK 004.932.72'1.02:656.121.025.2
IFAC 2.8; 1.1.8; 5.7

Original scientific paper

Previously, we introduced a passengers' counting algorithm in public rail transport. The main disadvantage of that algorithm is it lacks automatic event detection. In this article, we implement two automatic wavelet-based passengers counting algorithms. The new algorithms employ the spatial-domain Laplacian-of-Gaussian-based wavelet, and the frequency-domain applied Non-Linear Difference of Gaussians-based wavelet bandpass video scene filters 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 for detecting an object as seated is no longer needed as we now apply a boundary box tracker on the segmented moving objects' blobs. A scene map is combined with the wavelet-based methods and the boundary box for multi-camera object registration. We have developed a novel holistic geometrical approach for exploiting the scene map and the recorded video sequences from both cameras installed in each train coach to separate the detected objects and locate their positions on the scene map. We test all the algorithms with several video sequences recorded from the both cameras installed in each train coach. We compare the previously developed non-automatic passengers' counting algorithm with the two new automatic wavelet-based passengers' counting algorithms, and an additional spatial-domain automatic non-wavelet based Simple Mixture of Gaussian Models algorithm.

Key words: Video analytics, Event detection, Automatic passengers' seats counting, Wavelets, Laplacian-of-Gaussian, Non-Linear Difference of Gaussians, Simple Mixture of Gaussians, Illumination invariant, Frequency and spatial domain

Automatsko brojanje putnika u javnom željezničkom prijevozu uporabom waveleta. U prethodnim radovima uveli smo algoritam za brojanje putnika u javnom željezničkom prijevozu. Glavna manjkavost dosadašnjeg algoritma odsustvo je sustava za automatsko otkrivanje događaja. U ovom radu implementirali smo dva algoritma za automatsko brojanje putnika temeljena na waveletima. Novi algoritmi koriste LoG (Laplacian-of-Gaussian-based) wavelete u prostornoj domeni i pojasne filtre temeljene na waveletima nastalim na nelinearnim razlikama Gaussovih funkcija u frekvencijskoj domeni, pomoću kojih se izdvajaju značajke neosjetljive na razlike u osvjetljenju iz pojedine scene. Te značajke kombiniraju se u referentnu sliku koja prikazuje pozadinu scene. Ručna segmentacija scene u pravokutnike korištena u prethodnom algoritmu više nije potrebna jer se sada koristi automatsko praćenje rubova na segmentiranim objektima. Mapa scene kombinirana je s wavelet metodama i okvirom granica slike u svrhu registracije objekata pomoću više kamera. Razvili smo i novi cjeloviti geometrijski pristup koji koristi mapu scene i snimljeni videozapis iz dvije kamere postavljene u svakom vagonu vlaka pomoću kojeg možemo odvojiti detektirane objekte i locirati njihove položaje na mapi scene. Algoritmi su ispitani na nekoliko videosekvenci snimljenih s dvije kamere u vagonima. Usporedili smo ranije razvijene neautomatske algoritme za brojanje putnika s dva nova algoritma i s jednostavnim MoG algoritmom u prostornoj domeni.

Ključne riječi: analiza videozapisa, otkrivanje događaja, automatsko brojanje sjedalica, waveleti, LoG, nelinearna razlika Gaussovih funkcija, jednostavna MoG, neosjetljivost na promjene u rasvjeti, frekvencijska i prostorna domena

1 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 various scenarios' events [1–3]. In public transport, video surveillance cameras are being installed, and video analytics are becoming helpful. Their primary goal is to provide

additional security. However, the different conditions in vehicles turn the video analytics' task difficult.

While a lot of research recently has been conducted on the topic of video analytics, the number of publications for scenarios inside moving vehicles is quite limited. In [4], Milcent and Cai 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 and Gao [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 labour 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. A part of the work described in this paper, was presented in [11].

Next, in Section 2, we discuss briefly the previously developed non-automatic passengers' counting algorithm. In Section 3, we describe in details the two wavelet-based scene segmentation methods. In Section 4, we discuss about the Simple Mixture of Gaussian Models scene segmentation algorithm. In Section 5, we describe a novel scene map with a holistic geometrical approach for multi-camera objects' separation and registration. In Section 6, the bounding box tracking algorithm is given. In Section 7, the recorded results for the two wavelet-based algorithms, and the simple mixture of models algorithm for automatic passengers' counting are analyzed and discussed in comparison with the recorded results for the non-automatic passengers' counting algorithm. A brief subsection with indications on each algorithm's performance and computational complexity is included, too. Section 8 concludes and future research work is suggested.

2 NON-AUTOMATIC PASSENGERS' 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 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 able to be detected. For this purpose, rectangular regions are defined manually at the positions of the seats. Those rectangles are further subdivided in manually defined tiles. It is worth mentioning, that a major drawback in this algorithm is that Camera 1 (CAM1, see Fig. 1(a)) can process only half of the passengers' coach and Camera 2 (CAM2, see Fig. 1(b)) can process the other half of the passengers' coach. This is due to the limitations of drawing manually the rectangles and tiles with respect to the perspective ratio of the passengers' coach. 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. Figure 2(a) shows an example of rectangular regions being defined manually at the positions of the seats, and Fig. 2(b) shows the manually defined tiles inside those rectangles.

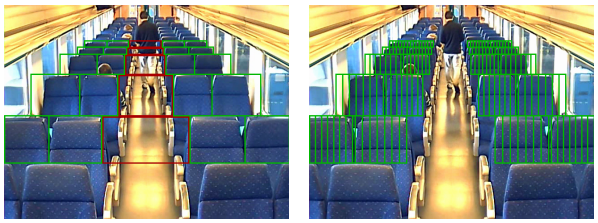
3 WAVELET-BASED AUTOMATIC PASSENGERS' COUNTING

3.1 Laplacian of Gaussian

In our previously described algorithm, we combined an edge detection algorithm 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 decided to work with a Laplacian of Gaussian (LoG) wavelet-based algorithm for efficiently segmenting the scene against illumination changes.

(a) Camera 1 (b) Camera 2

Fig. 1. An example of a test video sequence recorded by two cameras installed in each train coach



(a) Manually defined rectangular regions (b) Sub-division of rectangles with manually defined tiles

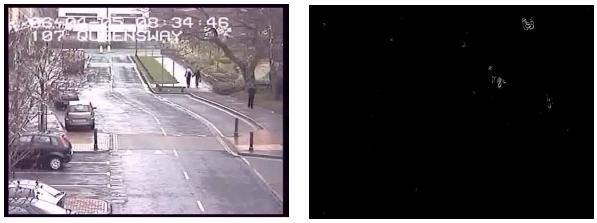
Fig. 2. Non-automatic passengers' counting algorithm

3.1.1 Illumination Invariance

To assess the illumination invariance of the LoG wavelet-based algorithm, we applied it on a test video sequence we have recorded outdoors. The selected test video sequence has a high degree of illumination changes and illumination sources, such as a curved wet surface road which has an increased reflectivity, curved wet vehicles metal surfaces which have an increased reflectivity, and shadows created from the trees on the road surface and on foreground objects (see Fig. 3(a)). Figure 3(b) shows LoG applied on a video frame of the sequence for segmenting the scene. It can be seen that LoG was able to be tuned appropriately for selecting only the discriminative features of the foreground scene objects and suppress the illumination changes and any high-frequency noise in the input frame scene.

3.1.2 Implementation

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 LoG operation $LoG(x, y)$, which is shown in (3). LoG can be shown that it acts as a bandpass filter. By selecting the right kernel dimensions, which for our test video sequences



(a) Frame from the illumination invariance video sequence (b) Laplacian of Gaussian wavelet-based scene segmentation applied on a video frame we used to demonstrate its illumination invariance properties

Fig. 3. Illumination invariance for Laplacian of Gaussian

was found to be a 7x7 size kernel, the background noise can be filtered out almost completely, while maintaining the edges.

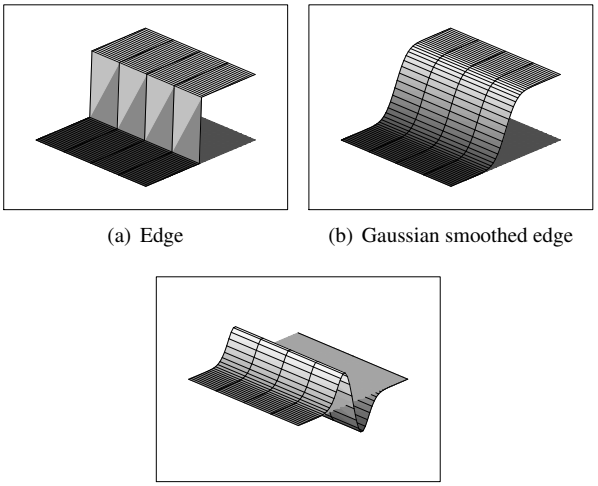
$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{1}$$

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

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

The applied LoG operation on the image edges, produces positive values at the one side of the edge and negative values at the other side (see Fig. 4). 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 LoG transformed (spatial domain) current frame is subtracted from the previous LoG transformed frame to obtain the moving edges in the current frame.

Figure 5 shows the results of the LoG operator applied on the test video sequence shown in Fig. 1. It can be seen that we were able to tune the kernel dimensions appropriately, so that LoG is able to be tuned for selecting only the discriminative pass-band mid-frequency features of the moving passengers and stop lower cutoff frequencies of illumination changes effects and any upper cutoff high frequency noise in the scene. Now, after segmenting the scene and detecting the foreground objects we can apply a tracking bounding box in the scene (see Section 6).



(a) Edge (b) Gaussian smoothed edge

(c) Laplacian of Gaussian edge detection

Fig. 4. Laplacian of Gaussian operator applied to an edge; the zero-plane is also plotted

3.2 Non-Linear Difference of Gaussians

By subtracting two Gaussian concentric kernels with different standard deviation values for a limited time duration, it forms a new kernel [12], which has an average value of zero, and it becomes 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 can be shown that is given by:

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

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) convolutioned image with the two Gaussian kernels. It can be found that DoG filter forms a type of band-pass filter with lower and upper cut-off frequencies set by the two Gaussian kernels. By tuning with appropriate values the standard deviation parameters σ_1 and σ_2, DoG filter is able to select the discriminative pass-band mid-frequency features as the foreground scene objects, and stop the low-frequency illumination changes effects and any high-frequency noise in the input image scene [13]. It is proven that DoG filter approximates best the Laplacian ∇^2 operator (or the two-dimensional second directional derivative of the Gaussian kernels ∇^2 for creating a narrow band-pass differential operator [13]) when the ratio of the two standard deviations σ_1/σ_2 is equal to 1.6.

(a) Camera 1 (b) Camera 2

Fig. 5. Laplacian of Gaussian wavelet applied on the test video sequences

The following observations can be made on DoG filter operation: (a) 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:

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

3.2.1 Illumination Invariance

As we did for LoG, in order to assess the NL-DoG wavelet-based filter's illumination invariance properties, we decided to apply it on the same, as before, test video sequence recorded outdoors which has a high degree of illumination changes and illumination sources. Figure 6(a) shows a video frame from the recorded test video sequence, and Fig. 6(b) shows NL-DoG applied on that video frame for segmenting the scene. It can be seen that NL-DoG was able to be tuned for selecting only the discriminative pass-band mid-frequency features of the foreground scene objects and suppress the low-frequency illumination changes effects and any high-frequency noise in the input frame scene.

3.2.2 Implementation

Though the NL-DoG filter is formulated in the spatial (time) domain, for calculating its coefficients when applied on the input images we use a Fast Fourier Transform (FFT)

(a) Frame from the illumination invariance video sequence

(b) Non-linear Difference of Gaussians wavelet-based scene segmentation applied on a video frame we used to demonstrate its illumination invariance properties

Fig. 6. Illumination invariance for Non-linear Difference of Gaussians

operation and, then, calculate them in frequency domain. Hence, (4) becomes:

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

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:

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

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.

Figure 7 shows the NL-DoG filter implementation for scene segmentation of foreground objects in the automatic passengers' counting algorithm we have developed. We used the previously described in details algorithm Time Intervals with Memory (TIME) for composing the reference background frame [15]. 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. NL-DoG filter is applied on each selected frame in the frequency domain. Then, NL-DoG filter's coefficients are computed (frequency domain) for each selected frame, and all the NL-DoG filter transformed frames are averaged to compose the background reference frame. In Fig. 7, assume for example that the test video sequence we used has a total duration of 35 seconds with 13 frames-per-second (fps),

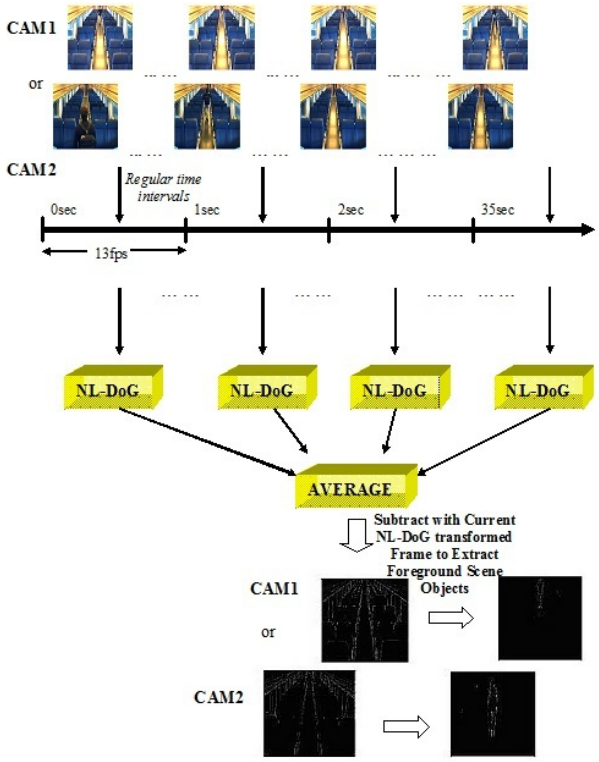
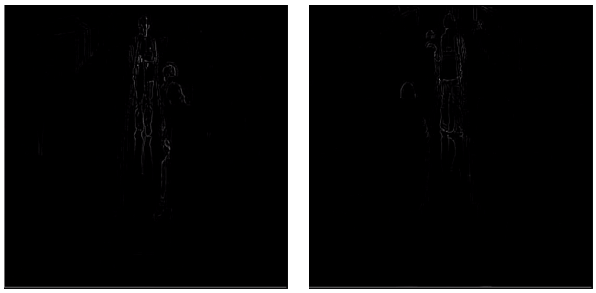


Fig. 7. Foreground scene objects extraction in the test video sequences recorded from Camera 1 and 2, installed in the passengers' train coaches, using TIME algorithm and NL-DoG wavelet-based filter

then we select one frame per fixed time intervals each second i.e. in total 35 frames were selected for composing the reference background frame. We apply NL-DoG filter in the frequency domain on the current test video sequence frame, and we subtract the composed background reference frame for extracting the foreground scene objects [16]. Figure 8 shows the results of the NL-DoG (TIME) wavelet-based filter applied on the test video sequence shown in Fig. 1. It is seen that NL-DoG wavelet-based filter has successfully selected all the foreground objects' features and suppressed the background scene features. Notice that NL-DoG has allowed a more uniform distribution of energy around the closed regions of the input frame scene. Thus, in the detected foreground objects finer details can be observed in comparison to the LoG where the energy seems to be concentrated around the edges of the objects. Now, as for LoG, after segmenting the scene and detecting the foreground objects we can apply a tracking bounding box in the scene (see Section 6).



(a) Camera 1 (b) Camera 2

Fig. 8. Non-linear difference of Gaussians applied on the test video sequences

(a) Frame from the illumination invariance video sequence

(b) Simple Mixture of Models non-wavelet based scene segmentation applied on a video frame we used to test its illumination invariance properties

Fig. 9. Illumination invariance for Simple Mixture of Models

4 SIMPLE MIXTURE OF MODELS

4.1 Illumination Invariance

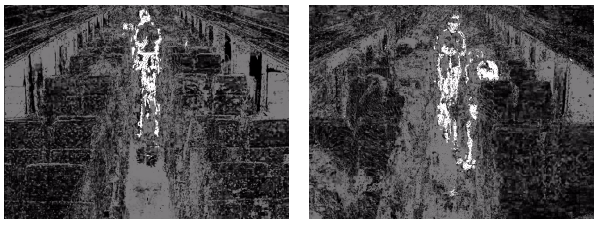
As we did for LoG and NL-DoG wavelet-based filters, to assess the simple mixture of models algorithm's illumination invariance properties, we applied it on the same, as before, test video sequence recorded outdoors which has a high degree of illumination changes and illumination sources. Figure 9(a) shows a video frame from the recorded test video sequence, and Fig. 9(b) shows simple mixture of models algorithm applied on that video frame for segmenting the scene. It can be seen from the segmented scene that the pedestrians on the pavement are being detected, while also the two pedestrians walking and a vehicle moving at the far back of the frame are also being detected. However, in overall the simple mixture of models algorithm is not performing as well as LoG and NL-DoG wavelet-based filters, since for example background features are detected and not adequately suppressed, such as the moving tree at the left hand side of the video sequence.

4.2 Implementation

Mixture of Gaussians Model (MGM) is one of the most popular background subtraction techniques, which can handle highly complex, multi-modal scenes with difficult situations like moving trees and bushes, clutter, noise, and permanent changes of the background. However, although MGM gives good results in many video surveillance applications, the use of the Gaussian models and the update scheme are complex.

To overcome the complexity of the traditional MGM, a simple mixture of models technique (SMM) is proposed by Poppe et al. [17]. The SMM models consist of an average, an upper and lower threshold, a maximum difference with the last background value, and an illumination allowance based on Skellam parameters. In many cases, only performing temporal background subtraction is insufficient, so SMM is extended with spatial information, i.e., fast edge-based image segmentation, to improve the detection results. The experimental results in [17] show that this advanced MGM method is more robust than 'standard' MGM and more recent techniques, resulting in less false positives and negatives. This is also the reason why SMM is selected as one of the non-wavelet based BG subtraction methods in our evaluation. For more detailed information on SMM, the reader can be referred to the original work.

Figure 10 shows the results of SMM algorithm when we applied it on the test video sequence shown in Fig. 1. In overall, the SMM algorithm is not performing as well as LoG and NL-DoG wavelet-based filters, since the resulted segmented scene contains significant amount of background noise. Though SMM algorithm has detected the foreground moving passengers it was not able to suppress dynamic illumination changes effects, such as passengers' shadows on the train coach floor and passengers reflections on different areas in the input scene. This can be explained due to the static nature of Skellam parameters which cannot be tuned for adapting to dynamically changing environments, such as the tested video sequences.



(a) Camera 1 (b) Camera 2

Fig. 10. Simple Mixture of Gaussian Models applied on the test video sequences

5 SCENE MAP: A HOLISTIC GEOMETRICAL APPROACH

We have applied a novel scene map [18] for multi-camera objects separation and registration. The scene map is independent of the actual train coach settings. We define separate monitoring zones in the scene as being, FAR-LEFT, FAR-RIGHT, MID-LEFT, MID-RIGHT, NEAR-LEFT, NEAR-RIGHT (see Fig. 11). In contrast to the manually defined rectangles and tiles of the non-automatic passengers' counting algorithm, here there is no explicit pixel counting for separating the scene into the different event monitoring zones. The detected objects can be classified as seated when they move to any of the –LEFT or –RIGHT zones of the map where the passengers seats are located.

Inspired by stereo vision [19, 20] methods and human vision cognitive processes [21] we use a novel holistic geometrical approach for separating the detected objects and positioning them in the scene map. In a similar fashion to GPS triangulation techniques [22] and to simultaneous localisation and mapping (SLAM) techniques [23, 24] used in robotics we exploit, first, the novel scene map we have developed and, second, the multi-camera view from both of the installed cameras in each train coach. Figure 11(a) and Fig. 11(b) show an example where we have used the novel holistic geometrical approach for separating the detected passengers and registering their position on the scene map.

On Fig. 11(a) (FRAME CAM1) and Fig. 11(b) (FRAME CAM2) assume the fixed co-ordinates of CAM1 to be $X_{C1} = M/2$, $Y_{C1} = 0$ and the fixed co-ordinates of CAM2 to be $X_{C2} = M/2$, $Y_{C2} = N$ for an $[M \times N] = [640 \times 480]$ resolution video frame. If we can extract from our algorithm the co-ordinates of the detected object (X_{Cobj}, Y_{Cobj}) and we already know the fixed co-ordinates of CAM1 (X_{C1}, Y_{C1}), then for each detected object we can calculate on Fig. 11(a) (FRAME CAM1) for Y-axis:

$$K^{obj}_{Y-CAM1} = \left| \frac{Y_{Cobj}}{N} - \frac{Y_{C1}}{N} \right|, \qquad (8)$$

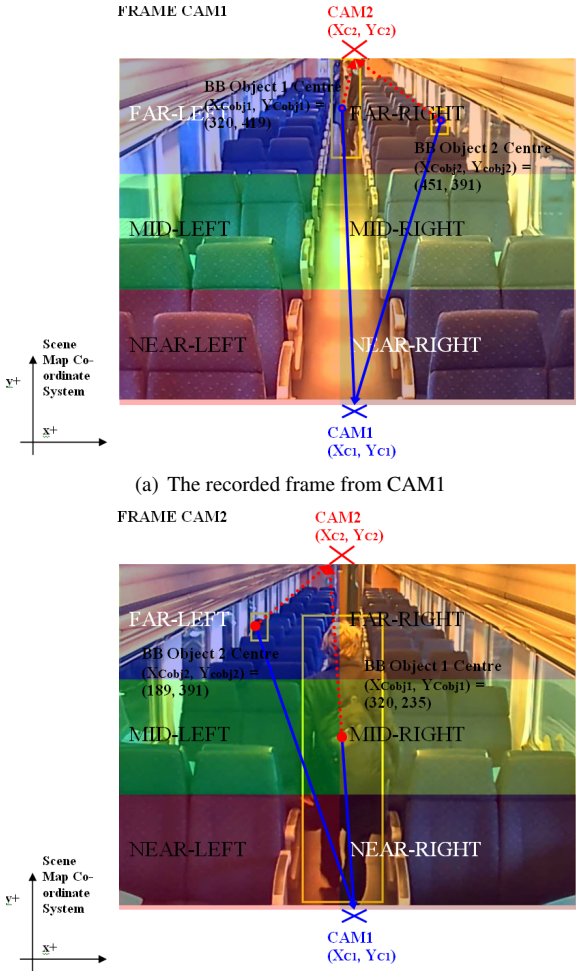
where K^{obj}_{Y-CAM1} is the unit normalised absolute distance for Y-axis of a detected object's Y-axis co-ordinate to the fixed Y-axis co-ordinate of CAM1.

$$K^{obj}_{Y-CAM2} = \left| \frac{Y_{Cobj}}{N} - \frac{Y_{C2}}{N} \right|, \qquad (9)$$

where K^{obj}_{Y-CAM2} is the unit normalised absolute distance for Y-axis of a detected object's Y-axis co-ordinate to the fixed Y-axis co-ordinate of CAM2.

Similarly, for X-axis we can calculate:

$$K^{obj}_{X-CAM1} = \left| \frac{X_{Cobj}}{M} - \frac{X_{C1}}{M} \right|, \qquad (10)$$



(a) The recorded frame from CAM1

(b) The recorded frame from CAM2

Fig. 11. Novel Scene Map with event monitoring zones shown as FAR-LEFT, FAR-RIGHT, MID-LEFT, MID-RIGHT, NEAR-LEFT, NEAR-RIGHT. We used a holistic geometrical approach for object separation and registration

where K^{obj}_{X-CAM1} is the unit normalised absolute distance for X-axis of a detected object's X-axis co-ordinate to the fixed X-axis co-ordinate of CAM1.

$$K^{obj}_{X-CAM2} = \left| \frac{X_{Cobj}}{M} - \frac{X_{C2}}{M} \right|, \qquad (11)$$

where K^{obj}_{X-CAM2} is the unit normalised absolute distance for X-axis of a detected object's X-axis co-ordinate to the fixed X-axis co-ordinate of CAM2.

Then, assuming that CAM1 and CAM2 have the same angle of view and the same view field width, the following conditions must be satisfied:

$$K^{obj}_{Y-CAM1} + K^{obj}_{Y-CAM2} = 1, \qquad (12)$$

$$K^{obj}_{X-CAM1} = K^{obj}_{X-CAM2}, \quad (13)$$

$$K^{obj}_{X-CAM1} = L^{obj}_{X-CAM1}, \quad (14)$$

where L^{obj}_{X-CAM1} is the unit normalised absolute distance for X-axis of a detected object's X-axis co-ordinate to the fixed X-axis co-ordinate of CAM1 for FRAME CAM2 (see Fig. 11(b)).

Eqns. (12), (13), (14) above can be proven by replacing with the co-ordinates values the equations' variables that are uniquely satisfied by the detected object's co-ordinates and not by any other detected object in the scene of FRAME CAM1.

Similarly, if we can extract from our algorithm the co-ordinates of the detected object (X_{Cobj}, Y_{Cobj}) and we already know the fixed co-ordinates of CAM1 (X_{C1}, Y_{C1}), then for each detected object we can write eqns. (8) and (9) on Fig. 11 (b) (FRAME CAM2) for Y-axis, and eqns. (10) and (11) on Fig. 11 (b) for X-axis. Then, assuming that CAM1 and CAM2 have the same angle of view and the same view field width, the following conditions must be satisfied:

$$L^{obj}_{Y-CAM1} + L^{obj}_{Y-CAM2} = 1, \quad (15)$$

$$L^{obj}_{X-CAM1} = L^{obj}_{X-CAM2}, \quad (16)$$

$$L^{obj}_{X-CAM1} = K^{obj}_{X-CAM1}, \quad (17)$$

where L^{obj}_{Y-CAM1} and L^{obj}_{Y-CAM2} are the unit normalised absolute distances for Y-axis of a detected object's Y-axis co-ordinate to the fixed Y-axis co-ordinate of CAM1 and of CAM2, respectively.

Eqns. (15), (16), (17) above can be proven by replacing with the co-ordinates values the equations' variables that are uniquely satisfied by the detected object's co-ordinates and not by any other detected object in the scene of FRAME CAM2.

Also, it can be proven that eqns. (12), (13), (14) for FRAME CAM1 and eqns. (15), (16), (17) for FRAME CAM2 are simultaneously satisfied only by Object 1 (linear algebraic Gauss system solution).

Therefore, from eqns. (12), (13), (14) and (15), (16), (17) we were able to distinguish between the different detected foreground objects and register their positions on the scene map for Frame CAM1 and Frame CAM2.

We can now extract, if we need to, metadata where the position of each detected object is shown on the scene map with their co-ordinates inside an event monitoring zone for FRAME CAM1 and FRAME CAM2, e.g. detected object 1 is in FAR-RIGHT zone for FRAME CAM1 and in NEAR-LEFT zone for FRAME CAM2.

6 BOUNDING BOX TRACKING

To achieve automatic passengers' counting in our developed algorithms, we implemented a Bounding Box (BB) tracking method for object classification. The segmented scenes by applying LoG (spatial domain) and NL-DoG (frequency domain) wavelet-based filters, and SMM (spatial domain) non-wavelet based algorithm with the detected foreground objects are passed to the BB tracker. First, a threshold value is used for the LoG, NL-DoG and SMM segmented scenes. Then, we use morphological closure and opening of the detected foreground objects. The produced objects' blobs are compared with those detected in the previous frame. If there is detected a similar object blob in the previous frame, BB is matched and passed on to the next step. In the next step, BBs are classified 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. In effect, certain BBs are recognised and others rejected by the tracker. The recognised BBs are used in the final step with the scene map and for multi-camera object separation and position registration (see Fig. 12(a) and Fig. 12(b)).

(a) Recognized bounding boxes for CAM1 (b) Recognized bounding boxes for CAM2

Fig. 12. Bounding box tracker

7 EVALUATION

7.1 Test Setup

We evaluated the different algorithms on video sequences 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 those sequences. An example of the recorded video sequences is shown on Fig. 1(a) for CAM1 and on Fig. 1(a) for CAM2. For all the test video sequences we have preprocessed them for clearing out any duplicate patterns of frames created during the acquisition and storing stage of the sequences. The used sequences are available at `http://multimedialab.elis.ugent.be/users/pdpotter/public_transport/`.

7.2 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 (TNR), and accuracy are calculated as follows:

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{18}$$

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{19}$$

$$\text{TNR} = \frac{TN}{TN + FP}, \tag{20}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}. \tag{21}$$

7.3 Simulation Results

We have formulated into several tables the results we have taken from the conducted tests with the video sequences. Table 1 summarizes the recorded performance metrics of LoG and NL-DoG wavelet-based filters and SMM algorithm, and the non-automatic passengers' counting algorithm for the conducted initial tests with *Short Cam1*, *Short Cam2*, *Long Cam1* and *Long Cam2* video sequences. *Short* sequence has a duration length of 35 seconds. *Long* sequence has a duration of 4 minutes approximately. For the non-automatic passengers' counting algorithm we show on Table 1 only one value instead of two, one for each camera. That is due to the limitations of that algorithm where CAM1 can be used to process only half of the passengers' coach and CAM2 can be used to process the other half of the passengers' coach. Thus, the value shown is the combined one for the two cameras.

Table 2 summarizes the results we have taken for LoG and NL-DoG wavelet-based filters and SMM algorithm, and the non-automatic passengers' counting algorithm. For those tests now we have 4 different video sequences, two ones recorded by CAM1 and two others recorded by CAM2. Thus, *Long 1 Cam1* and *Long 1 Cam2* have a duration length of approximately 4 minutes, and *Long 2 Cam1* and *Long 2 Cam2* have a duration length of approximately 2.50 minutes. Again, for the non-automatic passengers' counting algorithm we show on Table 2 only one value instead of two, one for each camera.

Table 1. Performance Metrics Values for NL-DoG and LoG wavelet-based automatic passengers' counting algorithms, for SMM non-wavelet automatic passengers' counting algorithm, and for non-automatic passengers' counting algorithm (Short Cam1, Short Cam2, Long Cam1 and Long Cam2 video sequences)

	Precision	Recall	TNR	Accuracy
Previous approach				
Short	1	0.8325	1	0.8839
Long	0.6366	0.9940	0.0274	0.6379
Laplacian of Gaussian				
Short cam1	0.9918	0.9032	0.9816	0.9258
Short cam2	1	0.8846	1	0.9162
Long cam1	0.9260	0.8468	0.1653	0.7957
Long cam2	0.9706	0.7819	0.3429	0.7667
Non-linear difference of Gaussians				
Short cam1	1	0.5146	1	0.603
Short cam2	1	0.5529	1	0.9679
Long cam1	0.9533	0.6641	0.2446	0.6467
Long cam2	0.8570	0.8867	0.0866	0.7752
Simple mixture of models				
Short cam1	1	0.5326	1	0.6204
Short cam2	1	0.5142	1	0.6015
Long cam1	0.9944	0.6002	0.7373	0.6020
Long cam2	0.9926	0.6301	0.6788	0.6311

On Table 3 we show the results we have taken for LoG and NL-DoG wavelet-based filters and SMM algorithm, but for only one of the sequences, namely the *Long 1 Cam2*. Here, we have shown the performance metrics values for each event monitoring zone separately (FAR-LEFT, FAR-RIGHT, MID-LEFT, MID-RIGHT, NEAR-LEFT, NEAR-RIGHT). Obviously, there are no values shown for the non-automatic passengers' counting algorithm since it does not use the novel scene map. In Appendix at the end of this article we have included Table 4, Table 5, and Table 6 where we have shown the results from the separate event monitoring zones taken for LoG and NL-DoG wavelet-based filters and SMM algorithm with the rest of the test video sequences *Long 1 Cam1*, *Long 2 Cam1*, and *Long 2 Cam2*.

From Table 1, Table 2 and Table 3 (and the rest of the tables in Appendix) the taken results for Precision and Accuracy have proven that the automatic wavelet-based passengers' counting algorithms and the automatic non-wavelet passengers' counting algorithm perform better than the non-automatic passengers' counting algorithm. In effect, the automatic passengers' counting algorithms are able to produce a higher number of TPs than the non-automatic passengers' counting algorithm. From the taken Recall performance metric values, it can be found that, in overall,

Table 2. Performance Metrics Values for NL-DoG and LoG wavelet-based automatic passengers' counting algorithms, for SMM non-wavelet automatic passengers' counting algorithm, and for non-automatic passengers' counting algorithm (Long 1 Cam1 and Long 1 Cam2, Long 2 Cam1 and Long 2 Cam2 video sequences)

	Precision	Recall	TNR	Accuracy
Previous approach				
Long 1	0.6135	0.9879	0.1232	0.6241
Long 2	0.8296	0.7975	0.3820	0.7111
Laplacian of Gaussian				
Long 1 cam1	0.9328	0.8646	0.3252	0.8190
Long 1 cam2	0.8859	0.8620	0.0964	0.7782
Long 2 cam1	0.9150	0.8118	0.3056	0.7621
Long 2 cam2	0.8372	0.8976	0.1065	0.7683
Non-linear difference of Gaussians				
Long 1 cam1	0.9807	0.6120	0.6377	0.6128
Long 1 cam2	0.9615	0.8278	0.2557	0.8034
Long 2 cam1	1.0000	0.5000	1.0000	0.5175
Long 2 cam2	0.8660	0.7462	0.0402	0.6703
Simple mixture of models				
Long 1 cam1	0.9712	0.6591	0.5310	0.6540
Long 1 cam2	0.9161	0.8559	0.1306	0.7959
Long 2 cam1	0.8981	0.6145	0.2658	0.5842
Long 2 cam2	0.9899	0.5844	0.8009	0.5907

Table 3. Performance Metrics Values for NL-DoG and LoG wavelet-based automatic passengers' counting algorithms, and for SMM non-wavelet automatic passengers' counting algorithm: Long 1 Cam2 video sequence with separate scene map event monitoring zones

	Precision	Recall	TNR	Accuracy
Laplacian of Gaussian				
Far-left	0.4142	0.9939	0.0486	0.4301
Far-right	0.9618	0.6087	0.7646	0.6232
Mid-left	0.9676	0.7791	0.9905	0.9340
Mid-right	0.2006	1.0000	0.3657	0.4528
Near-left	1.0000	0.5000	1.0000	0.5792
Near-right	0.0000	-	0.9665	0.9665
Non-linear difference of Gaussians				
Far-left	0.5433	0.9164	0.0371	0.5256
Far-right	1.0000	0.5000	1.0000	0.5315
Mid-left	0.2829	0.6864	0.3159	0.4205
Mid-right	0.1999	1.0000	0.1825	0.3211
Near-left	0.9997	0.5400	0.9993	0.6177
Near-right	-	-	1.0000	1.0000
Simple mixture of models				
Far-left	0.9870	0.5901	0.9799	0.6987
Far-right	0.9456	0.7540	0.6146	0.7399
Mid-left	0.9895	0.5024	0.9970	0.8184
Mid-right	0.3719	0.6367	0.6637	0.6573
Near-left	0.5241	0.9109	0.2848	0.5751
Near-right	-	-	1.0000	1.0000

for most of the test video sequences the non-automatic passengers' counting algorithm exhibits a higher value than the automatic (wavelet-based or non-wavelet) passengers' counting algorithms. Specifically, focusing on the *Long Cam1* and *Long Cam2*, and on *Long 1 Cam1* and *Long 1 Cam2* sequences, and if we take the average value of CAM1 and CAM2, then we get approximately 0.81 and 0.86 for LoG, 0.78 and 0.72 for NL-DoG, and 0.62 and 0.76 for SMM. Thus, it means that the non-automatic passengers' counting algorithm recognizes wrongly as a passenger less times than the automatic Passengers' counting algorithms. But that is expected, since the non-automatic algorithm applies a manual segmentation and recognition of the objects which obviously can outperform any automatic passengers' counting algorithm. That is the case for the True Negative Rate performance metric value, where focusing again on the *Long Cam1* and *Long Cam2*, and on *Long 1 Cam1* and *Long 1 Cam2* sequences, and if we take the average value of CAM1 and CAM2, then we get approximately 0.25 and 0.21 for LoG, 0.17 and 0.45 for NL-DoG, and 0.71 and 0.33 for SMM, but 0.03 and 0.12, smaller values, i.e. less TNs in comparison to FPs for the non-automatic passengers' counting algorithm.

Moreover, in overall, we can find that the wavelet-based NL-DoG and LoG automatic passengers' counting algorithms are performing better than the SMM non-wavelet automatic passengers' counting algorithm. In effect, focusing on the *Long Cam1* and *Long Cam2*, and on *Long 1 Cam1* and *Long 1 Cam2* sequences, and if we take the average Precision value of CAM1 and CAM2, then we get approximately 0.95 and 0.91 for LoG, and 0.90 and 0.97 for NL-DoG, but 0.99 and 0.94 for SMM. Similarly, if we take the average Accuracy value of CAM1 and CAM2, then we get approximately 0.78 and 0.80 for LoG, and 0.71 and 0.71 for NL-DoG, but 0.62 and 0.72 for SMM. This can be explained, as we saw earlier, due to the static nature of Skellam parameters which cannot be tuned for adapting to dynamically changing environments. In effect, SMM fails to completely suppress the background features when applied to the test video sequences.

Also, in overall, we can find that for all the video sequences NL-DoG wavelet-based automatic passengers' counting algorithm is performing slightly better than the LoG wavelet-based automatic passengers' counting algorithm. This can be explained due to NL-DoG allowing a more uniform distribution of energy around the closed regions of the input frame scene in comparison to LoG which, in turn, caused slightly more TPs for the NL-DoG.

Additionally, we can observe that both the wavelet-based LoG and NL-DoG automatic passengers' counting algorithm, and the non-wavelet SMM passengers' count-

ing algorithm perform less well for the video sequences *Long 2 Cam1* and *Long 2 Cam2* than the video sequences *Long 1 Cam1* and *Long 1 Cam2*. By observing carefully the test video sequences, this can be explained due to the higher scene complexity of *Long 2 Cam1* and *Long 2 Cam2*, i.e. more passengers moving in and out of the scene, and more reflections shown on the coach's surfaces. With an even closer observation, we can find that both the wavelet-based LoG and NL-DoG automatic passengers' counting algorithm, and the non-wavelet SMM passengers' counting algorithm perform less well for the video sequence *Long 2 Cam1* than the video sequence *Long 2 Cam2*. This can only be explained, after carefully examining the data, due to discrepancies on the frame rate created during the acquisition and storing stage of the sequence i.e. the frame rate was not kept constant throughout the complete duration of the video sequence which consequently caused non-deterministic errors in the scene segmentation stage.

7.4 Algorithm Performance and Complexity

We have used for running our simulations a dual-core CPU at 2.0 GHz with 4GB RAM. Non-wavelet SMM automatic passengers' counting algorithm was able to process the test video sequences at approximately 2.5773 frames per second, wavelet-based LoG automatic passengers' counting algorithm was able to process the test video sequences at approximately 10.2296 frames per second, and wavelet-based NL-DoG automatic passengers' counting algorithm at approximately 1.6129 frames per second. Thus, it is found that the wavelet-based LoG automatic passengers' counting algorithm is able to process the test video sequences at a higher rate. This can be explained due to the different complexity of each of the three algorithms. In effect, SMM algorithm's complexity lies on the computational operation of updating the background model each frame. NL-DoG algorithm's complexity lies on performing the FFT and IFFT computational operations on each test video sequence frame (IFFT is needed for applying the non-linear sigmoidal function). LoG algorithm's complexity lies on the Laplace Transform computational operation on each test video sequence frame. Therefore, NL-DoG algorithm has the higher complexity cost than the other two algorithms due to a higher total number of floating point operations needed which in turn increases the processing time of each frame.

8 CONCLUSION AND FUTURE WORK

This paper described two new automatic wavelet-based passengers' counting algorithms. One was based on LoG and we applied it on the spatial domain, and the other one was based on NL-DoG and we applied it on the frequency domain. We have shown that both are able to select illumination invariant features in the synthesis of their background reference frame. We used a novel scene map with a holistic geometrical approach for separating the detected objects and registering their positions in the scene for both cameras. We have developed a BB tracker for classifying the detected objects in the segmented scenes. We have extensively tested the wavelet-based LoG and NL-DoG and non-wavelet SMM automatic passengers' counting algorithms, and the previously developed non-automatic passengers' counting algorithm with several video sequences. From the recorded results, we have shown that both the wavelet-based and non-wavelet automatic passengers' counting algorithms are performing better than the non-automatic passengers' counting algorithm. We have shown that the wavelet-based LoG and NL-DoG automatic passengers' counting algorithms are performing, in overall, better than the non-wavelet SMM automatic passengers' counting algorithm. This can be explained due to the dynamically changing illumination conditions of the test video sequences which SMM is not able to adequately adapt for. Also, we have found that NL-DoG wavelet-based algorithm produced more TPs than LoG wavelet-based algorithm.

In future, we plan to continue experimenting with a more robust tracking mask to recognize multiple objects and to classify them into different categories [25, 26]. A more robust tracking mask can tackle the performance deterioration caused by objects' occlusions in the scene. Also, it worth performing a detailed computational complexity analysis of the wavelet-based algorithms of LoG and NL-DoG. Finally, we plan to exploit further the possible extraction of automatic forensics metadata for the video sequences with the application of the scene map.

APPENDIX A TABLES FOR SCENE MAP MONITORING ZONES OF SEQUENCES LONG 1 CAM1, LONG 2 CAM1, AND LONG 2 CAM2

Table 4. Performance Metrics Values for NL-DoG and LoG wavelet-based automatic passengers' counting algorithms, and for SMM non-wavelet automatic passengers' counting algorithm: Long 1 Cam1 video sequence with separate scene map event monitoring zones

	Precision	Recall	TNR	Accuracy
Laplacian of Gaussian				
Far-left	0.8199	0.9055	0.5172	0.7922
Far-right	0.9582	0.7840	0.6183	0.7703
Mid-left	1.0000	0.9920	1.0000	0.9944
Mid-right	-	-	1.0000	1.0000
Near-left	-	-	1.0000	1.0000
Near-right	-	-	1.0000	1.0000
Non-linear difference of Gaussians				
Far-left	0.8002	0.6289	0.4855	0.5953
Far-right	1.0000	0.5000	1.0000	0.5215
Mid-left	1.0000	0.9375	1.0000	0.9554
Mid-right	-	-	1.0000	1.0000
Near-left	-	-	1.0000	1.0000
Near-right	-	-	1.0000	1.0000
Simple mixture of models				
Far-left	0.9993	0.5678	0.9984	0.6558
Far-right	0.9859	0.5902	0.8592	0.6054
Mid-left	0.8161	0.9933	0.5880	0.8506
Mid-right	-	-	1.0000	1.0000
Near-left	-	-	1.0000	1.0000
Near-right	-	-	1.0000	1.0000

Table 5. Performance Metrics Values for NL-DoG and LoG wavelet-based automatic passengers' counting algorithms, and for SMM non-wavelet automatic passengers' counting algorithm: Long 2 Cam1 video sequence with separate scene map event monitoring zones

	Precision	Recall	TNR	Accuracy
Laplacian of Gaussian				
Far-left	0.9986	0.6805	0.9838	0.6972
Far-right	0.5476	0.9962	0.6381	0.7475
Mid-left	0.0000	-	0.9081	0.9081
Mid-right	0.0000	-	0.7938	0.7938
Near-left	-	-	1.0000	1.0000
Near-right	-	-	1.0000	1.0000
Non-linear difference of Gaussians				
Far-left	0.9646	0.5945	0.5703	0.5933
Far-right	1.0000	0.5000	1.0000	0.7419
Mid-left	0.0000	-	0.9693	0.9693
Mid-right	-	-	1.0000	1.0000
Near-left	-	-	1.0000	1.0000
Near-right	-	-	1.0000	1.0000
Simple mixture of models				
Far-left	0.9450	0.5574	0.4545	0.5516
Far-right	0.7045	0.9849	0.7764	0.8496
Mid-left	0.0000	-	0.9996	0.9996
Mid-right	-	-	1.0000	1.0000
Near-left	0.0000	-	0.8923	0.8923
Near-right	-	-	1.0000	1.0000

Table 6. Performance Metrics Values for NL-DoG and LoG wavelet-based automatic passengers' counting algorithms, and for SMM non-wavelet automatic passengers' counting algorithm: Long 2 Cam2 video sequence with separate scene map event monitoring zones

	Precision	Recall	TNR	Accuracy
Laplacian of Gaussian				
Far-left	0.7368	0.8339	0.7932	0.8099
Far-right	0.8610	0.7597	0.1451	0.6826
Mid-left	0.0000	-	0.9275	0.9275
Mid-right	0.0000	-	0.6508	0.6508
Near-left	-	-	1.0000	1.0000
Near-right	-	-	1.0000	1.0000
Non-linear difference of Gaussians				
Far-left	0.3999	0.9043	0.3280	0.5189
Far-right	1.0000	0.5854	1.0000	0.5997
Mid-left	0.0000	-	0.7649	0.7649
Mid-right	0.0000	-	0.7987	0.7987
Near-left	-	-	1.0000	1.0000
Near-right	-	-	1.0000	1.0000
Simple mixture of models				
Far-left	1.0000	0.5000	1.0000	0.7318
Far-right	1.0000	0.5858	1.0000	0.6000
Mid-left	-	-	1.0000	1.0000
Mid-right	0.0000	-	0.9795	0.9795
Near-left	0.0000	-	0.9758	0.9758
Near-right	0.0000	-	0.9233	0.9233

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. The authors would like to thank the NMBS (National Railway Company of Belgium) and Televic Rail NV for the setting in which the test sequences were recorded.

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, 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, 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, pp. 555–560, Mar. 2008.

[4] G. Milcent and Y. Cai, "Location Based Baggage Detection for Transit Vehicles," tech. rep., Carnegie Mellon University, 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*, pp. 19–25, IEE, Apr. 2004.

[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)*, pp. 431–42, Sept. 2007.

[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.*, p. 6 pp., June 2006.

[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*, pp. 59–66, July 2008.

[9] N. Liu and C. Gao, "Bi-directional passenger counting on crowded situation based on sequence color images," in *Adv. in Artif. Real.ity and Tele-Exist. 16th Int. Conf. Proc. (Lect. Notes in Comp. Sci. Vol. 4282)*, pp. 557–564, Nov. 2006.

[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*, pp. 1294–1298, Mar. 2011.

[11] P. De Potter, I. Kypraios, S. Verstockt, C. Poppe, and R. Van de Walle, "Automatic available seat counting in public transport using wavelets," in *Proc. ELMAR 2011, Zadar, Croatia*, pp. 79–83, Sept. 2011.

[12] D. Marr and E. Hildreth, "Theory of Edge-detection," *Proc. R. Soc. of Lond. Ser. B-Biol. Sci.*, vol. 207, pp. 187–217, Feb. 1980.

[13] O. Arandjelovic and R. Cippolla, "Achieving Illumination Invariance using Image Filters," in *Face Recognition* (K. Delac and M. Grgic, eds.), pp. 15–30, I-Tech Education and Publishing, July 2007.

[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, 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," in *Proc. EUSIPCO*, pp. 903–907, 2011.

[17] 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, pp. 107203-1–107203-13, Oct. 2008.

[18] I. Kypraios, "A multi-level alarm algorithm for recognising human movement and detecting abandoned baggages." tech. note, 2020Imaging ltd., Univ. of Sussex, U.K., June 2009. TN/2020/01090630.

[19] D. Lowe, "Three-dimensional object recognition from single two dimensional images," *Artif. Intell.*, vol. 31, no. 3, pp. 355–395, 1987.

[20] Y. Sumi, Y. Kawai, T. Yoshimi, and F. Tomita, "3d object recognition in cluttered environments by segment-based stereo vision," *Int. J. Comput. Vis.*, vol. 46, no. 1, pp. 5–23, 2002.

[21] G. Xu and Z. Zhang, *Epipolar Geometry in Stereo, Motion, and Object Recognition: A Unified Approach*. Kluwer Academic Publishers, 1996.

[22] B. Hofmann-Wellenhof, H. Lichtenegger, and J. Collins, *Global Positioning System: theory and practice*. Springer-Verlag, 5th revis. ed., 2001.

[23] M. Dissanayake, P. Newman, S. Clark, H. Durant-Whyte, and M. Csorba, "A solution to the simultaneous localisation and map building (slam) problem," *IEEE Trans on Robot. and Autom.*, vol. 17, no. 3, pp. 229–241, 2001.

[24] M. Walter and J. Leonard, "An experimental investigation of cooperative slam," in *Proc. of 5th IFAC Sympos. on Intell. Auton. Veh., Lisbon, Portugal*, July 2001.

[25] I. Kypraios, "Video analytics algorithms for smart IP camera for i-LIDS Sterile Zone scenario solution of multiple humans' movement monitoring," tech. rep., 2020Imaging ltd., Univ. of Sussex, U.K., Dec. 2008. TR/2020/04081202.

[26] I. Kypraios, "Video analytics algorithms for smart IP camera for i-LIDS Sterile Zone scenario solution of human synthetic movement monitoring," tech. rep., 2020Imaging ltd., Univ. of Sussex, U.K., Dec. 2008. TR/2020/02081202.

Pieterjan De Potter received the M.Sc. degree in engineering from Ghent University, Ghent, Belgium, in 2008. Between 2008 and 2010, he was involved in the IBBT (Interdisciplinary Institute for Broadband Technology) project Share4Health. He is currently pursuing the Ph.D. degree at Multimedia Lab, Ghent University - IBBT. His research interests include video analytics and Semantic Web technologies.

Ioannis Kypraios received his BEng in Computer Systems Engineering with first class honours. He has an MSc in Modern Digital communication Systems with distinction. He was awarded an EU Bursary and in 2004 he completed his Doctorate at University of Sussex in Fully-Robust Object Recognition Systems (Optics & Photonics). Dr. Kypraios's current research interests focus on signal, image and video processing with cross-discipline research conducted in Neuroinformatics for the study of human learning and memory, and Geomatics for climate change monitoring. He is the author of over 50 submitted and/or published papers in refereed international conferences, journals and technical reports, including recently a best-seller book in Object Recognition Systems (September 2012). Previously, he has organised and chaired a special session in "Advances in Image Processing" hosted by the Massachusetts Institute of Technology, USA. In August 2012 presented his work in human synthetic movement monitoring for an international symposium sponsored by the Russian Academy of Sciences. From 2004 to 2008 he was a Postdoctoral Researcher, and in 2008 Dr. Kypraios became a full-time Research Fellow (Research Assistant Professor) at University of Sussex. In January 2011, he became Senior Lecturer at Department of Engineering and IT, ICTM, London, and in April 2012 he was appointed as Head of the APEM Computing Labs (Remote Sensing and R&D Division) at Begbroke Science Park of University of Oxford, UK. Dr. Kypraios is a full member of several processional bodies including the IET, IEEE and OSA, and he is a registered FP7 Expert/Reviewer and a registered IEEE Expert Spokesperson. In February 2012, he was appointed as an IEEE Computational Intelligence Society (CIS) GOLD Subcommittee panel member.

Steven Verstockt received his Master degree in Informatics from Ghent University in 2003. At the end of 2007 he joined the ELIT Lab of the University College West-Flanders as a researcher. In 2008, he started a PhD on video fire analysis at the Multimedia Lab of the Department of Electronics and Information Systems of Ghent University - IBBT (Belgium). Since 2012 he works as a post-doctoral researcher in this lab. His research interests include video surveillance, computer vision and multi-sensor data fusion.

Chris Poppe received the Master degree in Industrial Sciences from KaHo Sint-Lieven, Belgium, in 2002 and received his Master degree in Computer Science from Ghent University, Belgium, in 2004. He joined the Multimedia Lab, Department of Electronics and Information Systems (ELIS), Interdisciplinary Institute for Broadband Technology (IBBT), where he obtained the Ph.D. degree in 2009. His research interests include video coding technologies, video analysis, and multimedia metadata extraction, processing and representation, with a strong focus on standardization processes.

Rik Van de Walle received his M.Sc. and PhD degrees in Engineering from Ghent University, Belgium in 1994 and 1998 respectively. After a visiting scholarship at the University of Arizona (Tucson, USA), he returned to Ghent University, where he became professor of multimedia systems and applications, and head of the Multimedia Lab. His current research interests include multimedia content delivery, presentation and archiving, coding and description of multimedia data, content adaptation, and interactive (mobile) multimedia applications.

AUTHORS' ADDRESSES

Pieterjan De Potter, M.Sc.
Department of Electronics and Information Systems,
Multimedia Lab,
Ghent University - IBBT,
Gaston Crommenlaan 8 bus 201, B-9050 Ledeberg-Ghent,
Belgium
email: pieterjan.depotter@ugent.be

Ioannis Kypraios, Ph.D.
Head of APEM Computing Lab,
Centre for Innovation & Enterprise,
Begbroke Science Park,
University of Oxford,
Begbroke Hill, Woodstock Road, Begbroke, Oxfordshire,
OX5 1PF
email: i.kypraios@ieee.org

Steven Verstockt, Ph.D.
Chris Poppe, Ph.D.
Prof. Rik Van de Walle, Ph.D.
Department of Electronics and Information Systems,
Multimedia Lab,
Ghent University - IBBT,
Gaston Crommenlaan 8 bus 201, B-9050 Ledeberg-Ghent,
Belgium
email: steven.verstockt@ugent.be, chris.poppe@ugent.be,
rik.vandewalle@ugent.be

Received: 2012-03-30
Accepted: 2012-09-08