Comparative evaluation of stationary foreground object detection algorithms based on background subtraction techniques

Álvaro Bayona, Juan Carlos SanMiguel, José M. Martínez
Video Processing and Understanding Lab
Escuela Politécnica Superior, Universidad Autónoma de Madrid, SPAIN
E-mail: {Alvaro.Bayona, Juancarlos.SanMiguel, JoseM.Martinez}@uam.es

Abstract

In several video surveillance applications, such as the detection of abandoned/stolen objects or parked vehicles, the detection of stationary foreground objects is a critical task. In the literature, many algorithms have been proposed that deal with the detection of stationary foreground objects, the majority of them based on background subtraction techniques. In this paper we discuss various stationary object detection approaches comparing them in typical surveillance scenarios (extracted from standard datasets). Firstly, the existing approaches based on background-subtraction are organized into categories. Then, a representative technique of each category is selected and described. Finally, a comparative evaluation using objective and subjective criteria is performed on video surveillance sequences selected from the PETS 2006 and i-LIDS for AVSS 2007 datasets, analyzing the advantages and drawbacks of each selected approach.

1. Introduction

Currently the automatic analysis of video surveillance sequences has become an area of very active research in response to the increasing demand of security issues in public areas [1][2]. Video surveillance systems aim to provide automatic analysis tools that may help the supervisor personnel in order to focus his/her attention when a dangerous or strange event takes place.

In this context, the detection of stationary objects is receiving a special attention because it is a critical analysis stage in applications like the detection of abandoned objects or parked vehicles frequently used in the surveillance of public areas. Additionally, the recognition of stationary objects in crowded unconstrained contexts is a challenging task. Issues related to occlusions (by moving or stationary objects), appearance variations (e.g., color composition, shape) as people move relatively to the camera, lighting changes, speed of the objects and the density of moving objects in the scene should be taken into account.

In the detection of stationary foreground objects, background-subtraction based approaches have become the most popular choice due to the common use of fixed cameras and the assumption that the illumination changes in the scene are gradual [3][4][10]. However, some works exists [5] that don't use this approach for analyzing static images.

In this paper we present a comparative evaluation of the stationary foreground object detection approaches based on background-subtraction[5][6]. Firstly, these approaches are hierarchically organized into different categories. Then, representative approaches of these categories are selected and discussed. Finally, the selected approaches are tested and compared identifying their advantages and drawbacks in two typical scenarios for video surveillance: the detection of abandoned objects and parked vehicles. This comparison is provided through an objective and a subjective evaluation of the selected approaches.

The remainder of this paper is structured as follows: section 2 describes the classification for background-subtraction based approaches, section 3 describes the ones selected to be compared, section 4 shows experimental results and section 5 closes the paper with some conclusions.

2. Classification of background-subtraction based methods for stationary object detection

In this section we describe the proposed classification for categorizing the stationary foreground detection approaches based on background-subtraction techniques (see Fig. 1). As most of the existing approaches incorporate some kind of tracking analysis in their system, we have decided to exclude the use of tracking from the criteria used in the classification.

Firstly, we have divided the existing approaches in two categories depending on their use of one or more background subtraction models.

Depending on the use of the foreground maps computed in the background subtraction analysis, one-model based approaches can be classified in:

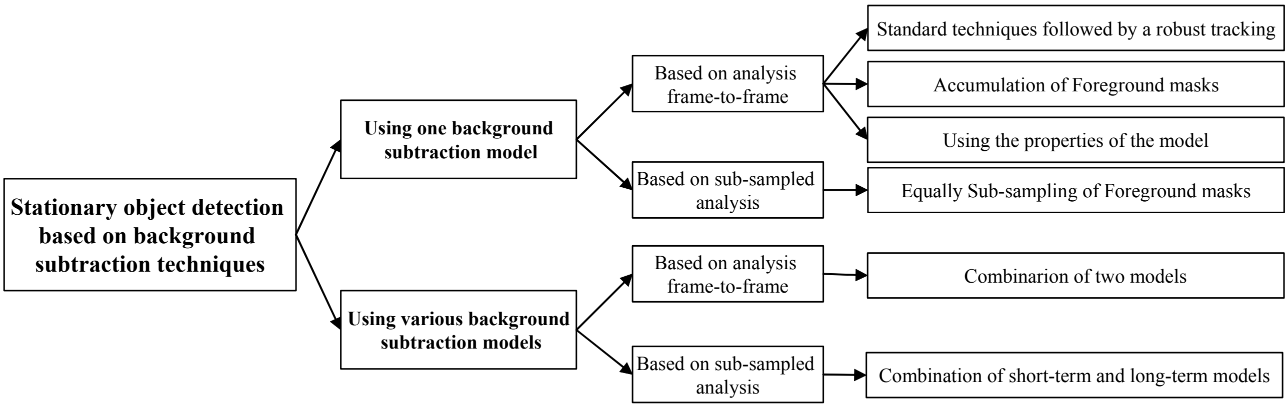

Figure 1: Classification of the background-subtraction based methods for stationary object detection

- *Based on frame-by-frame analysis.* This category describes the methods that employ typical background subtraction techniques followed by another type of analysis. Depending on this analysis stage we have the following approaches:
    - Based on the use of standard background techniques followed by another analysis stage (e.g., tracking) [3][10][18][19][20].
    - Based on the accumulation of foreground masks computed frame-by-frame [9][10][14].
    - Based on the properties of the background subtraction model used [11][12][13][15].
- *Based on a sub-sampled analysis*. These approaches try to detect stationary objects by analyzing the video sequence at different framerates [8][17].

Existing approaches combining two or more background subtraction models have been less investigated. However, a classification based on the processing framerate can be done as follows:

- *Based on frame-by-frame analysis.* In this category, we have methods that combine the different properties using two or more background subtraction techniques [21].
- *Based on a sub-sampled analysis*. These approaches detect stationary objects by analyzing the video sequence with various background subtraction methods at different framerates [7].

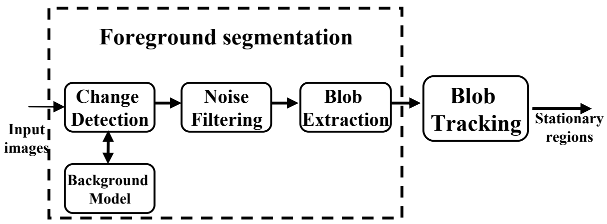

Figure 2: Stationary object detection procedure using [19]

3. Overview of selected approaches

In this section we describe the background-subtraction based approaches selected to be compared. We have chosen one representative approach for each previously described category attending to its implementation cost and detailed description in the papers studied.

3.1. One-model based

3.1.1 Based on frame-by-frame analysis

A. Based on the use of standard background techniques

As an example of this category, we have decided to implement the approach described in [19]. In this approach a typical background segmentation stage with a Gaussian Mixture Model (GMM) is proposed, followed by a blob tracking analysis stage. This tracking analysis is based on finding the correspondence between the blobs identified in two consecutive frames. Some rules, as colour, shape, distance or object size are used in this module to perform the tracking analysis. Fig. 2 depicts the processing scheme followed in the selected approach.

B. Based on the accumulation of foreground masks

As an example of this category, we have decided to implement the approach described in [14]. It is based on the accumulation of foreground masks to compute a confidence map to indicate the presence of stationary foreground objects.

In this algorithm, an intermediate image, $S(x,y)$, where each pixel indicates the confidence of the complexity image pixel belonging to a stationary object, is computed.

Initially, all the pixels of the confidence image are set to 0 being updated at every frame analysis. This update is based on the foreground masks obtained by previously applying a background subtraction stage. Two counter maps are calculated to update the confidence image: an increment counter $C(x,y)$, used when a pixel doesn't fit

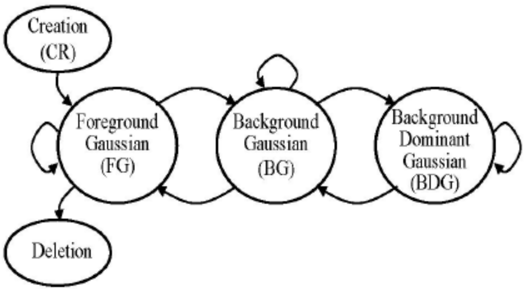

Figure 3: Transition states of distributions in GMM [12]

with the background (it is labelled as foreground), and a decrement counter *D(x,y),* used when a pixel fits with the background. The confidence image S is updated every frame pixel by pixel using these counters and depending on the defined *Framerate* of the video sequence and stationary object detection time *t*. Finally, *S* is thresholded to obtain a binary mask indicating the presence of stationary foreground objects.

C. Based on properties of the model

As an example of this category, we have decided to implement the approach described in [12]. It is based on the use of the GMM for detecting foreground objects and inspecting the properties of that model to detect stationary objects.

The stationary object detection is based on the observation of the transition states between the new Gaussian distributions created (for the new foreground pixels detected) and their transition to the dominant background state. Three Gaussians distributions are used in the GMM model resulting in the transition state diagram shown in Fig. 3. This approach describes a set of necessary conditions and corresponding observations on the transition state diagram to detect stationary objects imposing time stability, spatial stability and enough distribution weight constraints.

3.1.2 Based on a sub-sampled analysis

As an example of this category, we have selected the approach described in [8]. It is based on sampling the foreground-mask computed (see Fig 4). Firstly, a background subtraction stage based on modelling each pixel with a Gaussian distribution is performed. Additionally, a weight term is added for each pixel to take into account the gradual intensity change in image resolution or image deformations. Then, a number of sample foreground masks are taken from the last frames analyzed. The authors used 6 samples to determine the foreground mask (S) multiplying 6 binary foreground masks. Each active pixel of S (value equal to 1) indicates that a pixel has been foreground in the last 30 seconds and it presents a high probability of being stationary.

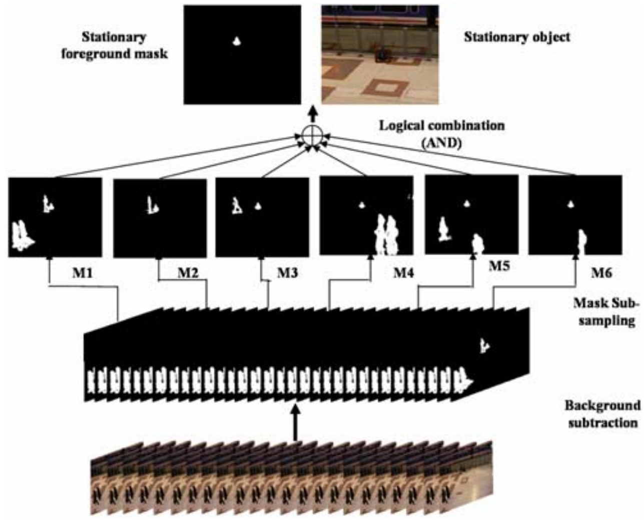

Figure 4: Foreground-mask subsampling procedure [8]

3.2. Two-model based

As an example of this category, we have decided to implement the approach described in [7]. In this method, a detection stage based on the application of two background subtraction methods at different framerate is proposed. The two models are based on the GMM employing one model for short-term detection (updating it every frame) and another for long-term detection (updating it every *n* frames). Short-term background is adapted faster and the scene changes are introduced more quickly on it. On the other hand, long-term background is adapted to the changes of the scene at a lower learning rate. Then, the foreground masks of the two models are computed at every frame and a combination of them is performed as shown in Fig. 5.

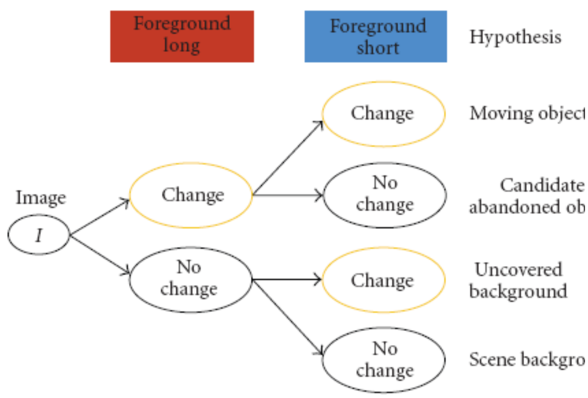

Figure 5: Combination of two background subtraction methods to detect stationary regions [7]

4. Performance evaluation comparison

In this section, experimental results of the selected approaches are presented and compared. The system has been implemented using the OpenCV image processing library (http://sourceforge.net/projects/opencv/). Tests were executed on a Pentium IV with a CPU frequency of 2.8 GHz and 1GB RAM. Additionally, we have decided to manually annotate the stationary regions corresponding to persons in the results obtained by the selected approaches in order to discard them. Then, stationary regions corresponding to objects have been evaluated with the ground-truth data provided in the datasets.

4.1. Experimental data

For the evaluation, we have selected two different types of sequences. Firstly, test sequences from the PETS 2006 dataset (available at http://pets2006.net/) have been selected as a simple scenario. This scenario presents lower stationary foreground extraction complexity, middle foreground object density and the speed of the scene objects is low. Secondly, test sequences from the i-LIDS dataset for AVSS2007 (available at http://www.elec.qmul.ac.uk/staffinfo/andrea/avss2007.html) have been selected as a complex scenario. This crowded scenario presents higher stationary foreground extraction complexity, higher foreground object density and the speed of the scene objects is variable ranging from low to high velocities.

4.2. Performance evaluation comparison metrics

The results of the selected approaches have been objectively and subjectively compared. Additionally, we have considered a foreground object as stationary if it remains static during 30 seconds.

As objective measures, we have selected the accuracy (in time) of the foreground object detection, the duration of the alarm/detection and Precision/Recall measures for the stationary foreground detection task.

For the subjective evaluation, we have selected 5 subjective measures. As the difficulty of the detection in crowded scenarios is determined by the occlusions between objects, moving object density and velocity, large perspective distortion, or the similarity in appearance of different people, we have decided to select a set of characteristics relative to these aspects (see Table 3). To evaluate them, we have ranged the characteristics between very low and very high.

4.3. Performance evaluation comparison

4.3.1 *Objective evaluation*

The results obtained from the experiments are summarized in Table 1 and 2 for, respectively, the simple and complex scenarios. Additionally, the precision and recall results of the stationary foreground detection task are presented in Table 3.

As it can be observed in Table 1, the overviewed approaches obtained similar results in the PETS2006 test sequences (simple scenario). In this scenario, stationary foreground objects are located in a low-dense area where the number of occlusions, lighting changes and moving objects is low. Additionally, colour dissimilarity between the stationary objects and the background allows a perfect identification by the background subtraction procedure.

On the other hand, as it can be observed in Table 2, results are more heterogeneous in complex (crowded) sequences. On these scenes, the presence of occlusions,

Table 1: *Objective evaluation of the selected approaches for the simple scenario*

PETS2006 Sequences	Ground Truth		Approach 1[19]		Approach 2[14]		Approach 3[12]		Approach 4[8]		Approach 5[7]	
	Start Time	Alarm Duration	Start Time	Alarm Duration	Start Time	Alarm Duration	Start Time	Alarm Duration	Start Time	Alarm Duration	Start Time	Alarm Duration
S1_T1_C3	1:52	0:08	1:52	0:08	1:52	0:08	1:55	0:06	1:52	0:08	1:52	0:08
S4_T1_C3	1:43	0:19	1:44	0:18	1:43	0:19	1:45	0:18	1:43	0:19	1:43	0:19
S5_T1_C3	1:26	0:26	1:26	0:26	1:26	0:26	1:22	0:30	1:26	0:26	1:26	0:26

Table 2: *Objective evaluation of the selected approaches for the complex scenario*

AVSS2007 Sequences	Ground Truth		Approach 1[19]		Approach 2[14]		Approach 3[12]		Approach 4[8]		Approach 5[7]	
	Start Time	Alarm Duration	Start Time	Alarm Duration	Start Time	Alarm Duration	Start Time	Alarm Duration	Start Time	Alarm Duration	Start Time	Alarm Duration
AB_Easy	2:05	0:36	2:03	0:37	2:06	0:37	2:14	0:38	2:05	0:36	2:06	0:37
AB_Medium	1:54	0:37	1:56	0:34	1:57	0:41	2:07	0:38	1:55	0:36	1:52	0:37
AB_Hard	2:11	0:56	2:12	0:54	2:17	0:59	2:27	0:57	2:11	0:56	2:15	0:50

Table 3: *Precision (P) and Recall (R) results for the stationary object detection task*

Approach	PETS2006 Sequences		AVSS2007 Sequences	
	P	R	P	R
Approach 1 [19]	0.05	1	0.01	1
Approach 2 [14]	0.6	1	0.1	1
Approach 3 [12]	0.5	1	0.03	1
Approach 4 [8]	0.75	1	0.33	1
Approach 5[7]	0.37	1	0.05	1

moving objects, and moving people is higher than in simple sequences. Approaches based on sub-sampling or accumulation (like [8][14]) obtain better results than methods based on simple background segmentation and tracking .

In Table 3, we can observe that all the annotated events are detected for both defined scenarios (Recall equals to 1). These results can be achieved by selecting the optimum parameters of each approach for each scenario. On the other hand, the Precision results show that the selected approaches perform the stationary foreground detection task with medium and very low accuracy for simple and complex scenarios respectively. This is due to the high amount of moving objects in the complex scenario. Approaches based on sub-sampling produce the best results due to the logical combination stage applied that eliminates most of the false positives.

Analyzing both scenarios, we can see that when increasing the number of mobile objects in the scene and the number of occlusions, results are less exact, but always with a minimal variance around ground truth results within a few seconds.

4.3.2 Subjective evaluation

The results of the subjective evaluation are reported in Table 4 for the simple and complex scenarios.

In general, analyzing the results obtained, we can observe that occlusions are completely removed from the final mask on sub-sampling approaches [8] and partially on the others [7][14]. In the simple scenario, all approaches (except [19]) present good results, but in complex scenarios (like the ones in the AVSS 2007 sequences), only [20] and [8] present good detection results.

Regarding the noise introduced, sub-sampling methods obtain better results because the noise (supposed to be statistically independent) is sub-sampled reducing its effect. In classic approaches, there is considerable noise in the mask and it should be removed in the following stages with different procedures (like noise filtering).

Computational cost is directly related with the number of background subtraction stages applied and the base technique used (GMM, KDE …). Approaches with light background-subtraction stages (like [8]) perform the detection faster than the others. The addition of additional analysis stages (like [19][14]) obviously increases the computational load. For example in [7], after the light stationary object detection, a light tracking stage is performed resulting in a low computational cost. Finally, approaches that apply various background subtraction stages add a low or high computational cost if, respectively, they analyze samples or the whole video sequence.

Parameter adjustment is difficult in approaches like [12] due to the need of fine tuning for inspecting the detection model. Nevertheless, more basic approaches (like [19]) do not present a high difficulty in the adjustment because slight errors in parameter settings can be corrected in the following analysis stages (using technologies like noise or shadow filtering). Sub-sampling approaches present medium difficulty because the sub-sampled time is the critical parameter of the scheme and it depends on the velocity of the objects and the framerate of the scene under analysis. Two model based approaches do not have so much parameterization problems except the time to update the models in sub-sampled schemes.

Table 4: *Subjective evaluation of the selected approaches*

Selected Approaches	Foreground Extraction accuracy	Tolerance to occlusions	Noise introduced	Computational Load	Parameter Adjustment Difficulty
Approach 1 [19]	Very Low	Very Low	Very High	Medium	Very Low
Approach 2 [14]	High	Medium	Low	Medium	High
Approach 3 [12]	Low	High	High	High	Very High
Approach 4 [8]	High	Very High	Very Low	Medium	Medium
Approach 5 [7]	Medium	High	Medium	Low	Low

5. Conclusions

This paper has presented a comparative evaluation of representative approaches based on background subtraction techniques for detecting stationary foreground objects. Firstly, the existing approaches have been classified into different categories. Then, representative approaches have been selected and described. Finally, an objective and subjective comparison has been performed.

Main conclusions of the study are the following. The results of the objective evaluation show that the detection of stationary foreground objects in simple scenarios is achieved with high accuracy in all the tested approaches. On the other hand, detection results for the complex scenario are more heterogeneous. Approaches based on sub-sampling schemes or accumulation of foreground masks assure the best results. In these type of scenarios the sub-sampling rate is a critical parameter (depending on the velocity of the moving objects) to determine the stationary objects. Subjective evaluation shows that sub-sampling based approaches obtain the best results on accuracy in the stationary foreground mask presenting a high tolerance to occlusions (frequently in complex scenes) and intermediate difficulty in parameters adjustment. In the case of complex scenarios, approaches based on standard background-subtraction techniques present the worst performance in all subjective measures due to the difficulty of analyzing the foreground masks in the following analysis stages (e.g., tracking). On the other hand, the difficulty in the adjustment of the parameters in these approaches is very low. Approaches based on the properties of the background-subtraction model used present low accuracy in the extracted mask due to the difficulty of the parameters adjustment phase. Concluding, for general-purpose stationary object detection, sub-sampling based approaches obtain the best results adding a low computational cost in the overall system.

6. Acknowledgments

This work has partially supported by the Cátedra UAM-Infoglobal ("Nuevas tecnologías de vídeo aplicadas a sistemas de video-seguridad"), the Spanish Administration agency CDTI (CENIT-VISION 2007-1007), by the Spanish Government (TEC2007-65400 SemanticVideo), by the Comunidad de Madrid (S-050/TIC-0223-ProMultiDis), by the Consejería de Educación of the Comunidad de Madrid, and by The European Social Fund.

7. References

[1] Plataniotis, K.N.; Regazzoni, C.S. “Visual-centric Surveillance Networks and Services”. IEEE Signal Processing Magazine, 22(2):12-15, 2005.

[2] Ferrando, S.; Gera, G.; Regazzoni, C. “Classification of Unattended and Stolen Objects in Video-Surveillance System”. Proc. of AVSS 2006, pp. 21-27.

[3] Beynon, M. “Detecting abandoned packages in a multi-camera video surveillance system”, Proc. of AVSS 2003, pp. 221–228.

[4] Li, L. “Statistical Modeling of Complex Backgrounds for Foreground Object Detection”, IEEE Trans. on Image Processing, 13 (11):1459-1472, 2004.

[5] Piccardi, M. “Background subtraction techniques: a review”, Proc. of SMC 2004, vol. 4, pp. 3099-3104.

[6] Spagnolo, P. et al. “An Abandoned/Removed Objects Detection Algorithm and Its Evaluation on PETS Datasets”. Proc. of AVSS 2006, pp. 17-21.

[7] Porikli, F.; Ivanov, Y.; Haga, T. “Robust Abandoned Object Detection Using Dual Foregrounds”, Journal on Advances in Signal Processing, art. 30, 11 pp., 2008.

[8] Liao,H-H.; Chang,J-Y.; Chen, L-G. “A localized Approach to abandoned luggage detection with Foreground –Mask sampling”, Proc. of AVSS 2008, pp. 132-139.

[9] Huwer, S.; Niemann, H. “Adaptive Change detection for Real-Time Surveillance Applications”, Proc. of Visual Surveillance 2000, pp. 37-46.

[10] Cheng, S.; Xingzhi Luo; Bhandarkar, S.M. “A Multiscale parametric background Model for Stationary Foreground Object Detection” , Proc. of Motion and Video Computing 2007, 8 pp.

[11] Miezianko, R.; Pokrajac, D. “Detecting and Recognizing Abandoned Objects in Crowded Environments”, Proc. of Computer Vision System 2008, pp. 241-250.

[12] Mathew, R.; Yu, Z.; Zhang, J. “Detecting new stable objects in surveillance video”, Proc. of Multimedia Signal Processing 2005, pp. 1-4.

[13] Guler, S.; Silverstein, J.A.; Pushee, I.H. “Stationary objects in multiple object tracking”, Proc. of AVSS 2007, pp. 248-253.

[14] Guler, S.; Farrow, K. “Abandoned Object detection in crowded places” , Proc. of PETS 2006, June 18-23.

[15] Bhargava, M.; Chen, C-C. Ryoo, M. S.; Aggarwal, J. K. “Detection of abandoned objects in crowded environments”, Proc. of AVSS 2007, pp. 271-276.

[16] Bird, N. “Real time, online detection of abandoned objects in public areas”, Proc. of Robotics and Automation 2006, pp. 3775-3780.

[17] Porikli, F. “Detection of temporarily static regions by processing video at different frame rates”, Proc. of AVSS 2007, pp. 236-241.

[18] Martínez, J.; Herrero, J.; Orrite, C. “Automatic Left luggage Detection and Tracking using a Multi-camera UKF”, Proc. of PETS 2006, pp 59-66.

[19] Stauffer, C; Grimson, W. E. L. “Adaptive background mixture models for real-time tracking,” Proc. of CVPR 1999, vol. 2, pp. 2246-2252.

[20] San Miguel J.C.; Martinez, J.M. “Robust unattended and stolen object detection by fusing simple algorithms”, Proc. of AVSS 2008, pp. 18-25.

[21] Tian, Y. “Robust and efficient foreground analysis for real-time video surveillance”, Proc. of CVPR 2005, pp. 1182-1187.