Evaluation of Background Subtraction Algorithms for Human Visual Surveillance

J. C. Neves
Instituto de Telecomunicações
University of Beira Interior
jcneves@ubi.pt

K. Wysoczanska
SOCIA: Soft Computing and Image Analysis Lab.
University of Beira Interior
wysoczanska.kamila@gmail.com

H. Proença
Instituto de Telecomunicações
University of Beira Interior
hugomcp@di.ubi.pt

***Abstract*—The fully automated surveillance of human beings remains an open problem, particularly for in-the-wild scenarios, i.e., for complex backgrounds and under uncontrolled lighting conditions. Background Subtraction (BGS) is typically the first phase of the processing chain of such type of systems and holds the feasibility of all the subsequent phases. Hence, it is particularly important to perceive the relative effectiveness of BGS, with respect to the kind of environment. This paper gives an objective evaluation of the state-of-the-art BGS algorithms on unconstrained outdoor environments. When compared to similar published works, the major novelties are two-fold: 1) the focus is put on scenes populated by human beings; and 2) an objective measure of the wildness of environments is proposed, that strongly correlates to BGS performance, and enables to perceive the algorithms' robustness with respect to the environment complexity. As main conclusions, we observed that the SOBS algorithm outperforms the remaining methods. Nevertheless, its performance leads to conclude that BGS in unconstrained environments is still an open problem.**

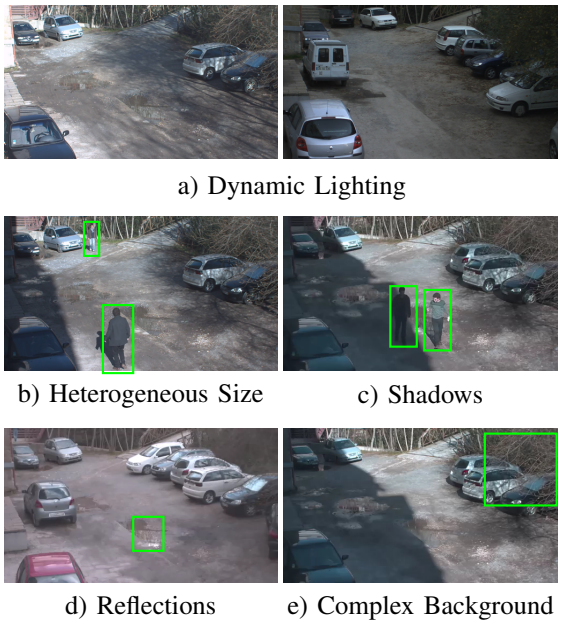
I. Introduction

Several attempts have been made toward the development of fully automated surveillance systems for action recognition / human identification purposes, but up to the moment no algorithm is robust enough to work in wild scenarios, i.e., under uncontrolled lighting conditions of outdoor environments.

Background Subtraction (BGS) is in the basis of the processing chain and provides support for all subsequent phases. Hence, the main goal of this paper is to evaluate the performance of the state-of-the-art BGS algorithms, with emphasis put on outdoor environments populated by human beings.

Even though previous evaluations of BGS techniques exist in literature, they have focused their interest in the post-processing performance [1], or in the general performance of BGS algorithms, regardless the type of object and the environment conditions [2]. The evaluation of BGS robustness to different kinds of degradation factors was addressed in the ChangeDetection Dataset [3], however surveillance scenarios were not given special attention and the methods were not exhaustively evaluated, i.e., only one parameter configuration was used, which may yield misleading results. The evaluation of BGS performance in surveillance videos was addressed in [4], but an artificial dataset was used which does not entirely capture the natural degradation factors, such as static shadows and reflections, Figure 1c) and 1d), respectively.

When compared to previously published works, this paper offers several discriminating features: 1) analysis of BGS methods performance in wild scenarios populated by human beings; 2) objective measurement of the environment *hardness*; and 3) analysis of performance variations with respect to parameter configurations.



a) Dynamic Lighting
b) Heterogeneous Size
c) Shadows
d) Reflections
e) Complex Background

Fig. 1. Typical challenges of BGS in-the-wild.

Figure 1 illustrates the major factors that determine the hardness of outdoor environments and decrease the average performance of BGS algorithms: a) dynamic lighting conditions mislead BGS methods by modifying the background/foreground distributions; b) heterogeneous size of subjects yield small foreground regions, prone to be erased by the post-processing step of BGS methods; c) the static shadows, cast by background objects, disturb the subjects appearance, whereas dynamic shadows, cast by foreground regions, disturb the appearance of the background ; d) reflections of the foreground objects are hardly ever identified as background; e) complex backgrounds require the methods ability to cope with periodic changes in the background.

In order to compare the performance of the different algorithms in a fair way, it is particularly important to perceive the effect of each of their parameters and to estimate the optimal configuration of each algorithm in the type of data used in our evaluation. Hence, we summarize each parameter used by the algorithms tested and report the levels of performance with respect to variations in each

978-1-4799-8996-6/15/$31.00

one.

Additionally, we also compare the relative performance of each BGS algorithm with respect to the hardness of environments, i.e., how their performance degrades as the data quality decreases, concluding not only about algorithms effectiveness but also about their robustness. Finally, we measure the individual impact of typical image degradation factors in the performance of BGS algorithms.

The rest of this paper is organised as follows: Section II provides a summary description of the BGS methods. Section III describes the proposed metric to quantify environment hardness and the experimental protocol used in our experiments. Section III also discusses the performance attained by the BGS methods in distinct environment conditions and in specific image degradation factors. Finally, conclusions are drawn in Section IV.

II. Background Subtraction: State-of-the-art

Despite most BGS methods rely on a background model, the strategy used to construct this model is the primary distinctive feature between them. Statistical analysis of the last N frames was one of the first strategies devised to model the background [5], [6], [7].

Gaussian-based methods assume that background chrominance is normally distributed. A single Gaussian [8] or a mixture of Gaussians [9] are used to encode the typical values of background.

In contrast to Gaussian-based methods, which assume pixel independence, the Eigenbackground method [10] takes advantage from the pixel correlations by building an eigenspace from a set of N background frames. However, it is not suitable to bootstrapping video sequences, since background frames are hardly ever found.

Alternatively, the Self-Organizing Background Subtraction (SOBS) algorithm [11] also uses neighbour correlation and its adaptive nature allows robustness to bootstrapping. The most likely HSV values of a background pixel are modelled by the weight vectors of a Self Organizing Map. Afterwards, the Spatially Coherent SOBS (SC-SOBS) [12] was proposed to generate spatially coherent foreground regions providing additional robustness against false detections.

The ViBe algorithm [13] distinguishes itself in the background modelling. Rather than infer a model for the typical values of each pixel, it samples the values of neighbour pixels.

III. Experiments and Results

A. Environment Wildness

Regarding human detection, the environment hardness is mainly dependent on the following factors: the number of subjects in the scene, the contrast between foreground and background, the lighting conditions, the static shadows of the background and the complex background.

In order to measure the hardness of an environment, we propose combining these factors in a single objective metric, hereinafter designed as wildness. Considering that the performance of human detection is directly related to environment wildness, we define the wildness of an environment as the miss rate of a person detector:

$$w = \frac{FN}{TP + FN}, \tag{1}$$

where FN and TP denote the number of false negatives and true positives yielded by the person detector.

In our experiments, we combined two detectors to achieve human detection: 1) the Viola-Jones detector [14], trained with human upper parts; 2) the HOG-based person detector [15]. The detectors were combined in the decision level and tested in a different set of videos (refer to section III-B for the dataset used).

Figure 2 illustrates how diverse scenarios are classified according to their wildness. Environments exempt of shadows, containing high contrast foreground regions and with few subjects are assigned with low wildness values. Contrarily, environments subjected to image degradation factors are classified as wild.

To provide additional support to the proposed metric we evaluated the correlation between BGS performance and the environment wildness (refer to section III-C for the detailed results).

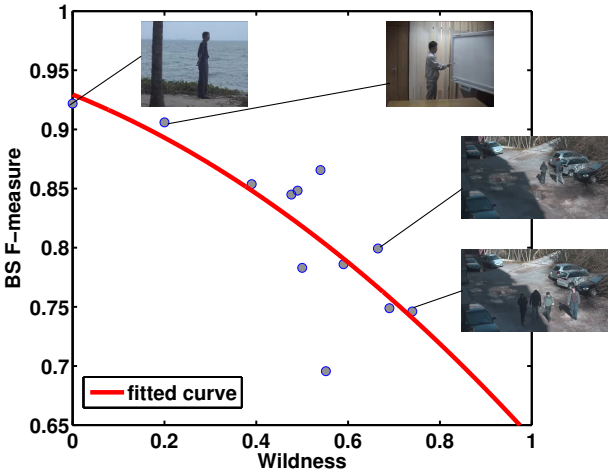


Fig. 2. Relation between the BGS performance and the wildness of different videos. Each point corresponds to the video wildness and the SOBS f-measure (BGS F-Measure). Illustrative frames of some videos emphasize the correlation between scenario conditions and the wildness measure.

B. Datasets and Covariates

In our comparative study, we collected a set videos captured in-the-wild, along with a set of videos commonly used in BGS evaluation [16], [11]. The description of both datasets used in our experiments is presented in Table I.

The wildness metric was used to separate the test videos in two datasets: unconstrained scenarios ($w > 0.5$) and controlled scenarios ($w \leq 0.5$). In this way, the videos were divided objectively rather than based on the dataset characteristics or visual observation.

Considering that the performance of each method is greatly dependent on parameter configuration, we performed an exhaustive search through the parameter space to find the optimal configuration of each method. Each

	SUVUBI	Perception [16]
Frame Rate	15	20
Resolution	512x288	320x256
Number of Frames	750	500-1200
Number of Videos	15	9
Conditions	Outdoor	Outdoor/Indoor
Dynamic Lighting	✓	
Complex Background	✓	✓
Shadow Interference	✓	
Low Contrast Foreground/Background	✓	

TABLE I. DETAILS OF THE DATASET CAPTURED FOR TESTING BS IN WILD ENVIRONMENTS.

Fig. 3. Example of the region of interest delimited for assessing the impact of a single image degradation factor.

BGS method was optimized independently in the different datasets, so that a correct performance comparison could be carried out.

Table II lists the complete set of parameters studied. The majority of the methods depend on a threshold (t) to separate background and foreground distributions. Median-based methods depend on a sampling rate (s) and a set of training frames t_f. Gaussian-based methods rely on a learning rate α and in n_g Gaussian distributions. ViBe uses a radius R a minimum number of background elements $\#$ and an update rate ϕ. The Eigenbackground uses n initial frames and m eigenvalues to model the background, whereas SOBS and SC-SOBS use different thresholds e and different learning rates c in the training phase and in online phase.

Apart from the BGS evaluation in distinct scenarios, we have also gauged the impact of image degradation factors in the performance of the BGS. In order to analyse each factor independently, we delimited a region of interest, containing only a specific factor. Figure 3 illustrates the region of interest gathered for the complex background factor. The performance of each algorithm in the region of interest is determined by the false positive rate: $FPR = \frac{FP}{FP+TN}$, where FP and TN denote the number of false positives and true negatives.

C. Performance Comparisons

Figure 4 presents the results obtained by the exhaustive search of the parameter space of each BGS method. Each point corresponds to precision and recall of a specific configuration. Different colours are used to illustrate the individual impact of a parameter when the remaining are fixed.

Figure 5 summarises the best performance attained by the described methods. The blue curves represent constant f-measure values and improve the visual perception of the overall performance of each method comparatively to the others.

Algorithms	Parameters
Frame Difference [5]	$p = (t)$
Adaptive Median [6]	$p = (t, s, t_f)$
Temporal Median [7]	$p = (t, s, b_f)$
EigenBackground [10]	$p = (t, n, m)$
Single Gaussian [8]	$p = (t, \alpha, t_f)$
MoG [9], [17]	$p = (t, \alpha, n_g)$
SOBS [11], [12]	$p = (e_1, e_2, c_1, c_2)$
ViBe [13]	$p = (R, \#, \phi)$

TABLE II. THE BGS ALGORITHMS EVALUATED IN THIS STUDY AND THE CORRESPONDING PARAMETERS.

Table III presents the performance comparison regarding different image degradation factors, as well as the results attained in unconstrained and controlled scenarios.

As expected, Frame Difference approach yielded the worst performance in both controlled and unconstrained scenarios. On the contrary, Adaptive Median proved to be more robust than Single Gaussian method, which can be explained by the adaptive nature of this approach, allowing the modelling of the dynamic changes of wild scenarios. The exhaustive search results corroborated this conclusion, showing that lower sampling rates improved the algorithm performance.

Temporal Median is based on the same principle of Adaptive Median, but it was more robust to ghosts, dynamic lighting conditions and complex background. Combining N sub-sampled frames with previous background models was the major reason for the Temporal Median achievements. The new background model encoded more recent information than the model of the Adaptive Median approach, and thus was more to resilient to changes in the background. The incomplete detection of the objects was the major drawback of this strategy.

When compared to the remaining algorithms, the Eigenbackground attained very poor results not only in dynamic lighting conditions, but also in unconstrained scenarios. Modelling the background by the first N frames was the primary cause for these results. The dynamic conditions of in-the-wild environments could not be encoded only by the initial frames, and thus adaptive approaches outperformed this method.

Regarding the image degradation factors, the performance of the Single Gaussian was considered reduced, comparatively to the remaining approaches. The reasons for these results were twofold: 1) complex backgrounds were not unimodal; 2) dynamic lighting conditions required fast adaptation.

Contrarily to Single Gaussian, the MoG approach assumes a multimodal distribution for each pixel, and consequently it attained good results in complex backgrounds and ghosts. The poor performance attained in dynamic light conditions was again justified by the reduced learning rate. With regard to the general performance of the MoG method, the excessive sensitivity to non-periodic background changes caused a plethora of false positives. Besides, the results of the exhaustive search showed that neither the threshold nor the learning rate could improve significantly the precision of the method.

Algorithms	Global performance (%)						Image degradation factors performance (%)				
	Controlled			Uncontrolled			DL	G	CB	SS	DS
	R	P	F	R	P	F	False Positive Rate (FPR)				
Frame Difference [5]	65.1	37.4	47.5	49.6	38.6	43.4	5.6	1.9	46.8	85.5	30.8
Adaptive Median [6]	81.9	90.4	85.9	80.5	73.4	76.8	12.1	54.9	10.2	83.7	95.5
Temporal Median [7]	86.1	90.0	88.0	81.3	71.5	76.0	2.3	4.8	0.4	99.4	86.2
Eigenbackground [10]	83.5	81.3	82.4	78.6	64.3	70.8	23.8	3.7	1.5	63.8	94.0
Single Gaussian [8]	76.3	85.5	80.6	76.8	68.1	72.2	16.6	57.5	14.9	57.1	91.0
MoG [9]	85.1	84.9	85.0	83.5	73.7	78.3	24.5	0.0	0.4	39.0	99.9
Improved MoG [17]	70.6	90.1	79.2	80.6	72.6	76.4	15.5	0.8	18.0	43.0	97.5
SOBS [11]	86.7	96.7	91.4	82.0	80.6	81.3	9.1	0.8	0.2	67.1	92.1
SC-SOBS [12]	85.1	96.9	90.6	80.9	80.8	80.9	7.2	0.9	0.1	83.4	89.7
VIBE [13]	80.9	93.5	0.867	69.6	82.0	0.753	7.38	0.84	0.6	92.5	86.1

TABLE III. SUMMARY OF THE PRECISION (P), RECALL (R) AND F-MEASURE (F) ATTAINED FOR EACH BGS ALGORITHM IN DIFFERENT ENVIRONMENTS. THE FALSE POSITIVE RATE (FPR) IS ALSO PRESENTED FOR THE FOLLOWING IMAGE DEGRADATION FACTORS: DYNAMIC LIGHTING (DL), GHOSTS (G), COMPLEX BACKGROUND (CB), STATIC SHADOW (SS) AND DYNAMIC SHADOW (DS).

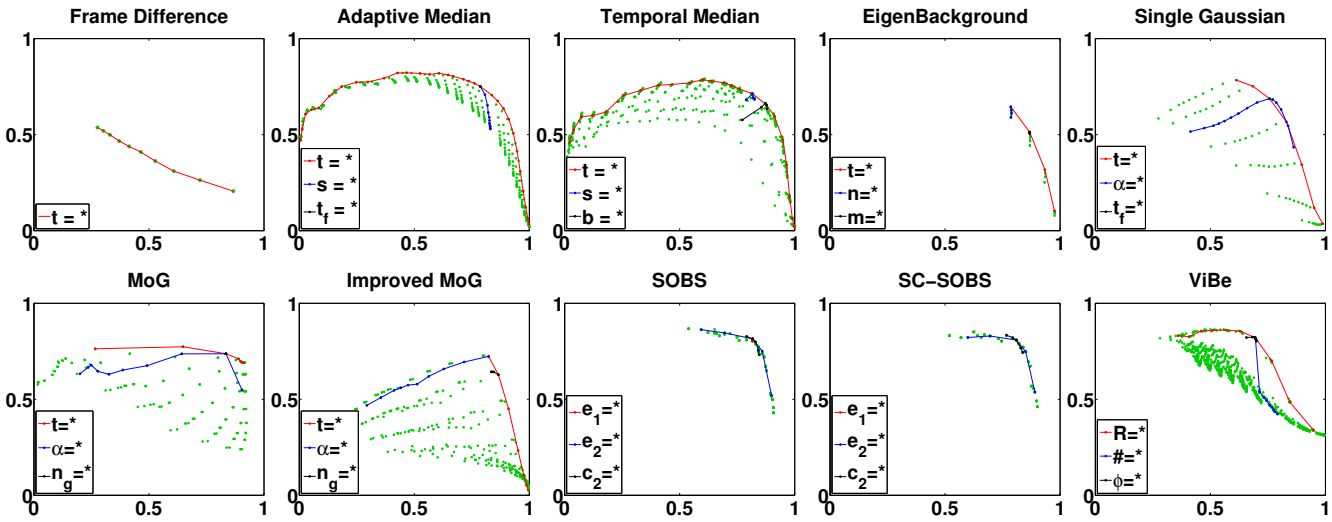


Fig. 4. Precision (y-axis) and recall (x-axis) obtained for each BGS method using different parameter configurations. Each green point denotes the performance achieved using a specific configuration, whereas each line represents the variation of a single parameter ($p = *$).

ViBe distinguished itself by its precision, which is mainly justified by the use of neighbour information in the background model, ensuring spatial consistency. However, this strategy also provides incorrect classification of object boundaries, which increases the false negative number.

Among the analysed methods, SOBS attained the best results. This approach was able to achieve good results in all the image degradation factors, maintaining an interesting overall performance. The use of a Self Organizing Map per pixel provided robustness to complex backgrounds, since the typical values of the multiple backgrounds were encoded in the different weight vectors. Besides, the relation between neighbour neurons boosted the adaptation to lighting changes.

Comparing to SOBS, the spatial coherence introduced by the SC-SOBS reduced the number of false detections. However, this improvement yielded a lower recall rate, and thus its general performance was worse than SOBS in wild scenarios.

With regard to the comparison between dynamic and controlled scenarios, we have determined the average performance of each method in the two environments. Figure 6 presents the average precision and recall of each method in controlled scenarios. For comparison purposes, each point was added a vector defined by:

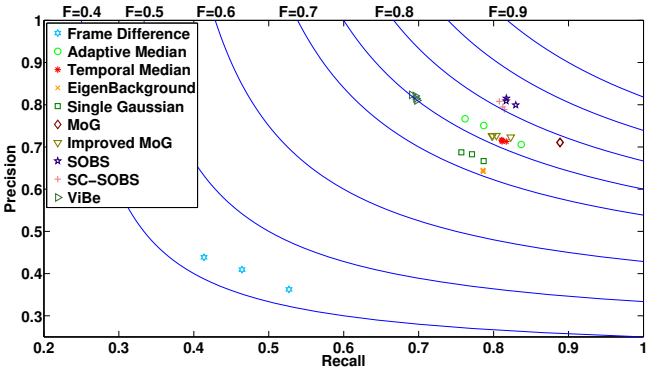


Fig. 5. The best three configurations obtained for each BGS method using an exhaustive search of the parameter space. Blue lines denote the set of points with constant f-measure (F).

$$\boldsymbol{v} = (r_u, p_u) - (r_c, p_c), \tag{2}$$

where R_u and R_c denote the average recall obtained in unconstrained and controlled scenarios, respectively. P_u and P_c denote the average precision obtained in unconstrained and controlled scenarios, respectively.

The results obtained evidence that BGS methods suffer

from performance degradation when addressing unconstrained scenarios. On average, each method degraded its f-measure in 11%. Besides, it was observed that in the majority of the methods the angle of $\boldsymbol{v}$ was approximately 102^o, implying that precision is more sensitive to environment conditions than recall. This can be easily explained by the impact of image degradation factors in BGS performance. Although they affect the detection of actual foreground regions, their effect on background regions is much more significant, and consequently the number of false positives increases more than the number of false negatives.

Additionally, we have also determined a relation between the BGS performance and the environment wildness. For this purpose, we determined the correlation between the SOBS performance and the wildness of an environment. The Pearson correlation between the two variables was determined to be -0.79, evidencing that BGS performance and wildness are inversely proportional. Furthermore, a quadratic relation was found between the f-measure of SOBS (BSF) and the wildness of a video (w):

$$BSF = -0.13w^2 - 0.16w + 0.93. \quad (3)$$

In short, the most important findings of our study are the following:

- BGS methods suffer from performance degradation in unconstrained scenarios when compared to controlled environments;
- Median-based methods adapt quickly to sudden changes in the scene, maintaining an acceptable recall rate. These methods have a good trade-off between the performance in image degradation factors and their general performance in wild scenarios;
- Although MoG has attained good performance in unconstrained scenarios, it is not adequate for highly dynamic environments containing non-periodic changes;
- ViBe has distinguished itself by its precision, however miss-detection of object parts represents its major drawback;
- By maintaining a good performance in the different image degradation factors and by attaining the best general performance, SOBS is the best method to address in-the-wild scenarios;
- In general, BGS methods are not robust to shadows. No algorithm has stood out in the dynamic shadows, whereas Gaussian-based methods attained the best performance in static shadows, mainly due to their high sensitivity to changes in the background;
- The best method (SOBS) attained a f-measure of approximately 81%, thereby we can conclude that BGS in-the-wild remains an open problem.

IV. Conclusions

In this paper, we presented a comparative analysis of the performance of the state-of-the-art BGS methods in wild scenarios. Additionally, we introduced an objective metric to classify the hardness of an environment, avoiding the subjective labelling of the test videos.

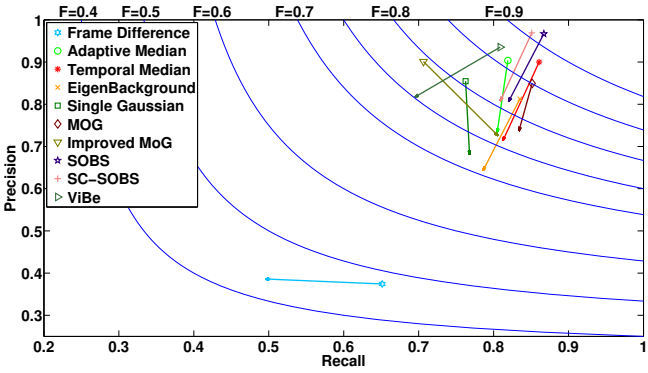


Fig. 6. Comparison between the average performance of BGS methods in distinct scenarios. Each point represents the average performance of a BGS method in controlled scenarios, and each vector represents the performance degradation in unconstrained scenarios. Blue lines denote the set of points with constant f-measure (F).

The obtained results evidenced that BGS methods degrade their performance in wild scenarios. Besides, we found that Median-based methods are adequate for highly dynamic scenarios, mainly due to their adaptation capacity. However, the miss-detection of some objects parts is the primary drawback of these methods.

The overall performance of each method stood out MoG, SC-SOBS and SOBS as the best ones to address unconstrained scenarios. However, MoG degraded its performance in environments with non-periodic changes and SC-SOBS attained similar results to SOBS with lower recall rates.

Although the obtained results can be considered satisfactory, the performance of SOBS (f-measure $\approx 81\%$) led us to conclude that BGS in unconstrained scenarios remains an open problem.

References

[1] D. Parks and S. Fels, "Evaluation of background subtraction algorithms with post-processing," in *IEEE Fifth International Conference on Advanced Video and Signal Based Surveillance*, Sept 2008, pp. 192–199.

[2] F. E. Baf, T. Bouwmans, and B. Vachon, "Comparison of background subtraction methods for a multimedia learning space," in *International Conference on Signal Processing and Multimedia - SIGMAP 2007*, July 2007, pp. 153–158.

[3] N. Goyette, P. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, "Changedetection.net: A new change detection benchmark dataset," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, June 2012, pp. 1–8.

[4] S. Brutzer, B. Hoferlin, and G. Heidemann, "Evaluation of background subtraction techniques for video surveillance," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2011, pp. 1937–1944.

[5] R. Jain and H.-H. Nagel, "On the analysis of accumulative difference pictures from image sequences of real world scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 206–214, April 1979.

[6] N. McFarlane and C. Schofield, "Segmentation and tracking of piglets in images," *Machine Vision and Applications*, vol. 8, no. 3, pp. 187–193, May 1995.

[7] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1337–1342, Oct 2003.

[8] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: real-time tracking of the human body," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 780–785, Jul 1997.

[9] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, Aug 1999, pp. 246–252.

[10] N. Oliver, B. Rosario, and A. Pentland, "A bayesian computer vision system for modeling human interactions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 831–843, Aug 2000.

[11] L. Maddalena and A. Petrosino, "A self-organizing approach to background subtraction for visual surveillance applications," *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1168–1177, July 2008.

[12] L. Maddalena and A. . Petrosino, "The sobs algorithm: What are the limits?" in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, June 2012, pp. 21–26.

[13] O. Barnich and M. Van Droogenbroeck, "Vibe: A universal background subtraction algorithm for video sequences," *IEEE Transactions on Image Processing*, vol. 20, no. 6, pp. 1709–1724, June 2011.

[14] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vision*, vol. 57, no. 2, pp. 137–154, May 2004.

[15] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, June 2005, pp. 886–893 vol. 1.

[16] L. Li, W. Huang, I.-H. Gu, and Q. Tian, "Statistical modeling of complex backgrounds for foreground object detection," *IEEE Transactions on Image Processing*, vol. 13, no. 11, pp. 1459–1472, Nov 2004.

[17] Z. Zivkovic, "Improved adaptive gaussian mixture model for background subtraction," in *Proceedings of the 17th International Conference on Pattern Recognition*, vol. 2, Aug 2004, pp. 28–31.