Long-term Stationary Object Detection based on spatio-temporal change detection

Diego Ortego, Juan C. SanMiguel, and José M. Martínez

***Abstract*—We present a block-wise approach to detect stationary objects based on spatio-temporal change detection. First, block candidates are extracted by filtering out consecutive blocks containing moving objects. Then, an online clustering approach groups similar blocks at each spatial location over time via statistical variation of pixel ratios. The stability changes are identified by analyzing the relationships between the most repeated clusters at regular sampling instants. Finally, stationary objects are detected as those stability changes that exceed an alarm time and have not been visualized before. Unlike previous approaches making use of Background Subtraction, the proposed approach does not require foreground segmentation and provides robustness to illumination changes, crowds and intermittent object motion. The experiments over an heterogeneous dataset demonstrate the ability of the proposed approach for short- and long-term operation while overcoming challenging issues.**

***Index Terms*—Stationary Object Detection, Long-term, Stability Changes, Online Clustering, Abandoned Object.**

I. Introduction

STATIONARY Object Detection (SOD) has recently experienced extensive research [1] due to its contribution to prevent terrorist attacks by detecting abandoned objects [2] and illegal parked vehicles [3]. SOD aims to detect the objects in the scene that remain stationary after previous motion. Typically, a Background Subtraction (BS) algorithm extracts the objects and SOD decides whether they are stationary or not [4]. However, current BS algorithms present many shortcomings to label foreground and background regions in real situations [5], thus highly determining the SOD accuracy.

Recent SOD approaches employ different strategies based on BS. Whilst temporal accumulation of foreground masks [6] is widely used, post-processing [7][8] and combination [9] of additional features are required to address BS limitations in presence of crowds and illumination challenges. Temporal sampling of foreground and motion masks enables operation in complex scenes [10], being the sample selection crucial for the detection accuracy. Dual BS approaches rely on fast and slow updated BS algorithms to identify the stationary objects [2][11]. However, BS failures require additional post-processing, such as edge features or fast-slow model interaction [12], which avoids detections of background objects that are removed from the scene. Other approaches take advantage of multilayer BS algorithms to model moving objects, stationary objects and background [13]. Nevertheless, the validation of the detected candidates via patch features [1] and edge features [14] is again needed to handle BS errors. Moreover, [3] detects parked vehicles over time using stable keypoints instead of BS. Many SOD challenges addressed in previous research are related to BS difficulties with illumination changes, crowds, intermittent object motion and required temporal adaptation. These aspects are pivotal to transfer SOD research to real situations, where long-term operation may be required. The increasing interest in long-term operation is reflected in recent works for abandoned object detection [1] and vehicle tracking [15], where BS limitations are addressed to reduce the high number of false alarms.

This letter proposes a SOD approach for long-term operation that has three main contributions. Firstly, it does not use BS to perform SOD, thus not being constrained to BS limitations. Secondly, the scene is modeled by an Online Block Clustering approach that describes the stationarity of the scene. The proposed approach is robust to illumination changes and quickly adapts to scene variations while identifying the stationary objects. Finally, few parameters are needed to operate with the proposed approach, unlike most of the state-of-the-art where handling stationary objects with BS introduces many parameters and thresholds. We validate the proposed approach for short-term and long-term scenarios, outperforming the state-of-the-art results.

This letter is organized as follows: Section II overviews the proposed approach whereas Section III and IV describe the clustering and the stationary detection. Section V presents the experiments. Finally, Section VI concludes this letter.

II. Overview

The proposed approach detects stationary objects without using BS (see Figure 1). A block-wise online clustering of the scene detects spatio-temporal stability changes at regular sampling instants. Those changes are exploited to identify stationary objects. Firstly, a *Block Division* stage decomposes each frame I_t into non-overlapping $N \times N$ blocks $B_t^{\mathbf{b}}$ at each instant t, where $\mathbf{b}$ denotes the block location. Secondly, an *Online Block Clustering* stage (see Section III) models each location $\mathbf{b}$ over time, updating a cluster partition $\mathcal{L}^{\mathbf{b}}$. This stage handles the temporal adaptation to scene changes, by assigning each incoming block $B_t^{\mathbf{b}}$ to one cluster of the partition $\mathcal{L}^{\mathbf{b}}$ or creating a new one. Only stationary blocks $B_t^{\mathbf{b}}$ (i.e., without motion with respect to $B_{t-1}^{\mathbf{b}}$) are analyzed at this stage. This clustering provides robustness against illumination changes by considering pixel ratios at block level which groups blocks even if their illumination has changed. Finally, a *Stationary Block Detection* stage (see Section IV) outputs

D. Ortego, J.C. SanMiguel and J.M. Martínez are with the TEC Department, Universidad Autónoma de Madrid (UAM), Spain, e-mail: {diego.ortego, juancarlos.sanmiguel, josem.martinez}@uam.es.

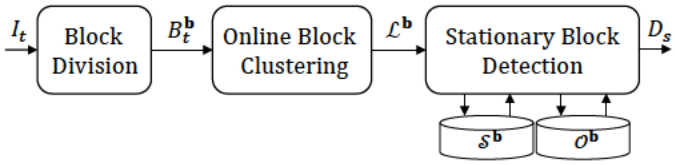

Figure 1. Block diagram of the proposed approach.

a result image D_s with stationary objects, where s defines the sampling instant each k frames. Data associated to the last stable cluster $\mathcal{S}^{\mathbf{b}}$, old stable clusters $\mathcal{O}^{\mathbf{b}}$ and the alarm time T is used to respectively detect the spatio-temporal stability changes, discard those changes caused by previously visualized clusters (i.e. empty scene or previous detections) and detect stationarity for changes longer than the alarm time. This last stage improves the state-of-the-art by reducing false alarms due to intermittent object motion and allowing to detect stationarity for objects not fully visible during T. Figure 2 presents an example of the scene analysis.

III. Online Block Clustering

Once each I_t is divided into non-overlapping blocks $B_t^{\mathbf{b}}$, the Online Block Clustering models the temporal scene evolution by grouping similar blocks over time into clusters. Thus, cluster partitions $\mathcal{L}^{\mathbf{b}} = \{C_q^{\mathbf{b}}\}_{q=1:|\mathcal{L}^{\mathbf{b}}|}$ are created for each block location $\mathbf{b}$, where $C_q^{\mathbf{b}}$ is the block representing each cluster and $|\cdot|$ denotes the cardinal. For a more readable notation, we omit the index $\mathbf{b}$ since the clustering operations apply to the same location.

As the target is to identify stationarity, blocks containing moving objects are not necessary. Therefore, matching between blocks from consecutive frames is first performed to discard each incoming B_t not matching B_{t-1}. For each B_t without motion, this stage determines the matches with existent clusters from $\mathcal{L}$. Each cluster models a spatio-temporal scene pattern and it is described by the first instant of visualization f_q, last instant of visualization l_q, repeatability w_q and the cluster representative C_q. To update $\mathcal{L}$, if no matching is found a new cluster with representative $C_{q'} = B_t$ is created, where $q' = |\mathcal{L}| + 1$. On the contrary, a match exists and an existent cluster is updated applying a cumulative moving average:

$$C_q = \frac{C_q^{match} \cdot w_q + B_t}{w_q + 1}, \tag{1}$$

where C_q^{match} is the cluster matching B_t and w_q is the block repeatibility computed as $w_q = f(C_q, B_{t-T:t})$. This function $f(\cdot)$ is computed online without keeping every B_t and counts each matching between C_q and B_t, also decreasing the w_q value in each sampling instant s to reduce the contribution of old visualizations. For example, given a sampling t, such decrease removes from the repeatability old visualizations summed before the instant $t - T$, i.e. older than the alarm time T from the current instant t, thus facilitating a fast adaptation for long-term operation. Hence, $B_{t-T:t}$ depicts that visualizations from $t - T$ to current instant t guide the computation of the w_q associated to C_q, operation that is performed online and without buffer. Occasionally, B_t can match different members of $\mathcal{L}$ and C_q^{match} is selected as:

$$C_q^{match} = \underset{\forall C_q \in \mathcal{C}}{argmax}\ f(C_q, B_{t-T:t}), \tag{2}$$

where $\mathcal{C} = \{C_q : C_q\ matches\ B_t\ \forall q\}$. Thus, the selected match with C_q, i.e. C_q^{match} to use in Eq. 1, has the highest repeatability w_q. Furthermore, in each sampling instant $\mathcal{L}$ is pruned to keep the z most visualized clusters in order to speedup the matching search.

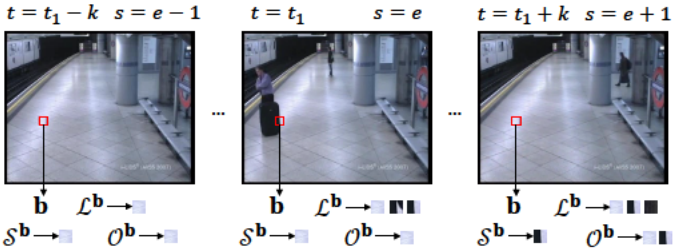

Figure 2. Example of the temporal analysis for a block location $\mathbf{b}$ where the stability is modified changing from the empty scene to a suitcase. $\mathcal{L}^{\mathbf{b}}$ keeps clusters from $\mathbf{b}$, while $\mathcal{S}^{\mathbf{b}}$ and $\mathcal{O}^{\mathbf{b}}$ keep, respectively, the information from the last stable cluster and old stable clusters. Additionally, the relation between the temporal index t and the sampling instant s is shown.

A. Matching metric

We determine the distance between two blocks B and B' based on pixel ratios, which are known to be robust against illumination for motion detection [16][17]. Inspired by these works, we define the ratio between two RGB pixels of B and B' as the maximum of their three-channel pixel ratio:

$$r_{max}(B(\mathbf{p}), B'(\mathbf{p})) = max\{r_i(B_i(\mathbf{p}), B'_i(\mathbf{p}))\}_{i=R,G,B}, \tag{3}$$

where $\mathbf{p}$ denotes a pixel location and r_i is the pixel ratio of each image channel i. The ratio of each channel is:

$$r_i(B_i(\mathbf{p}), B'_i(\mathbf{p})) = 1 - \frac{min\{B_i(\mathbf{p}), B'_i(\mathbf{p})\} + m}{max\{B_i(\mathbf{p}), B'_i(\mathbf{p})\} + m}, \tag{4}$$

being m a correction constant [16] to manage the ratio instability in low intensity values. Unlike [16], when comparing two pixel values we divide the minimum between the maximum value in order to obtain $r(\mathbf{p}) \in [0, 1]$. Thus, $r(\mathbf{p}) = 0\,(1)$ means maximum (minimum) pixel similarity.

To model the block as a whole, we use a feature vector $\mathbf{d}$ composed of mean μ and variance σ^2 of the pixel ratio:

$$\mu(B, B') = \frac{1}{|B|}\sum_{\mathbf{p}\in B} r_{max}(B(\mathbf{p}), B'(\mathbf{p})), \tag{5}$$

$$\sigma^2(B, B') = \frac{1}{|B| - 1}\sum_{\mathbf{p}\in B}\left(r_{max}(B(\mathbf{p}), B'(\mathbf{p})) - \mu(B, B')\right)^2. \tag{6}$$

Using mean and variance allows measuring, respectively, the intensity and heterogeneity of the variations between the blocks B and B'. The higher (lower) the intensity change is, the higher (lower) the $\mu(B, B')$ will be, while the higher (lower) the heterogeneity in the change is, the higher (lower)

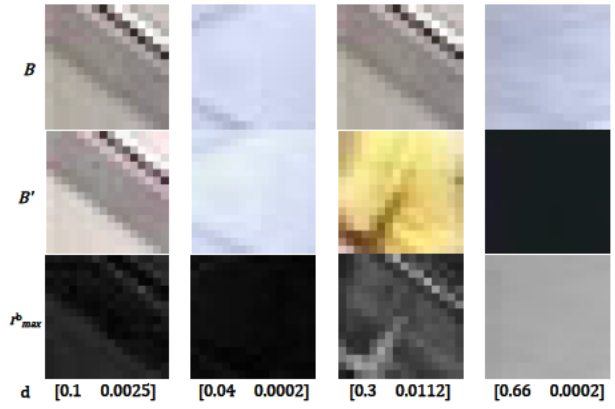

Figure 3. Ratio of blocks B and B', where $r^{\mathbf{b}}_{max}$ is the pixel ratio for every pixel $\mathbf{p} \in B$. The higher the difference, the brighter the pixel ratio. First and second columns are examples of equal blocks where $\mu(B, B')$ and $\sigma^2(B, B')$ have low values. Third and fourth columns are different blocks where $\mu(B, B')$ has a high value, while $\sigma^2(B, B')$ is high (low) in the third (fourth) column due to a heterogeneous (homogeneous) change.

$\sigma^2(B, B')$ will be. This behavior is depicted in Figure 3. The matching between B and B' is modeled by a pre-trained SVM based on the two-dimensional feature vector $\mathbf{d}$. Thus, for the Online Block Clustering, $\mu(B_t, B_{t-1})$ and $\sigma^2(B_t, B_{t-1})$ are computed to match B_t and B_{t-1} when discarding motion blocks, while $\mu(B_t, C_q)$ and $\sigma^2(B_t, C_q)$ are obtained to associate a non-moving B_t to any of the C_q members of $\mathcal{L}$.

IV. Stationary Block Detection

This stage analyzes stability in regular sampling instants (i.e. each k frames) to identify stationarity. To that end, data associated to old stable clusters $\mathcal{O}^{\mathbf{b}}$ and the last stable cluster $\mathcal{S}^{\mathbf{b}}$ in each block location is kept. The former contains the clusters generating stationary detections, i.e. $\mathcal{O}^{\mathbf{b}} = \{O^{\mathbf{b}}_h\}_{h=1:|\mathcal{O}^{\mathbf{b}}|}$, while the latter has the last stable cluster that either induced stationarity or was an old visualization.

In each sampling instant s, a sequence of operations is performed (summarized in Figure 4) to determine stationarity. First, the most stable cluster from $\mathcal{L}^{\mathbf{b}}$, $C^{\mathbf{b}}_s$, is obtained:

$$C^{\mathbf{b}}_s = \underset{C^{\mathbf{b}}_q \in \mathcal{L}^{\mathbf{b}} \forall q}{argmax}\ w^{\mathbf{b}}_q. \tag{7}$$

Subsequently, the occurrence of the spatio-temporal stability change is verified by comparing first visualization instant from $C^{\mathbf{b}}_s$ and $\mathcal{S}^{\mathbf{b}}$, i.e. $f^{\mathbf{b}}_s$ and $f^{\mathbf{b}}_*$:

$$\mathbb{S}^{\mathbf{b}}_s = \begin{cases} 1 & if \quad f^{\mathbf{b}}_s \neq f^{\mathbf{b}}_* \\ 0 & otherwise \end{cases}, \tag{8}$$

where $\mathbb{S}^{\mathbf{b}}_s$ denotes whether a stability change is occurring (1) or not (0). Note that the first instant of visualization is sufficient to verify equality between clusters, as it is an exclusive cluster footprint at each block location $\mathbf{b}$. Then, the buffer $\mathcal{O}^{\mathbf{b}}$ is consulted to determine if $C^{\mathbf{b}}_s$ was previously seen:

$$\mathbb{N}^{\mathbf{b}}_s = \begin{cases} 1 & if \quad C^{\mathbf{b}}_s \notin \mathcal{O}^{\mathbf{b}} \\ 0 & otherwise \end{cases}, \tag{9}$$

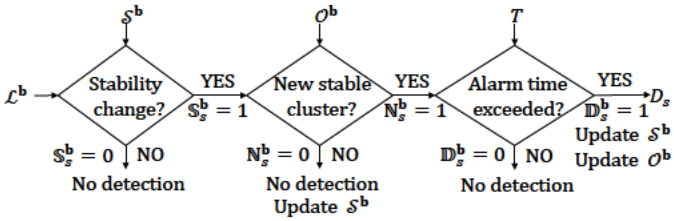

Figure 4. Stationary Block Detection. Sequence of operations to determine stationarity in a s sampling instant.

where $\mathbb{N}^{\mathbf{b}}_s$ determines whether $C^{\mathbf{b}}_s$ is a new stable cluster (1) or not (0). The matching measure from Section III is used for this task, i.e. $\mu(C^{\mathbf{b}}_s, O^{\mathbf{b}}_h)$ and $\sigma^2(C^{\mathbf{b}}_s, O^{\mathbf{b}}_h)$ are computed for every h to classify $C^{\mathbf{b}}_s$ as equal or different to any $O^{\mathbf{b}}_h$. Furthermore, in case of been and old cluster ($\mathbb{N}^{\mathbf{b}}_s = 0$), $\mathcal{S}^{\mathbf{b}}$ is updated with $C^{\mathbf{b}}_s$ and $f^{\mathbf{b}}_s$ to be the last stable cluster from $\mathbf{b}$. When $\mathbb{N}^{\mathbf{b}}_s = 1$, the stationary detection is performed as:

$$\mathbb{D}^{\mathbf{b}}_s = \begin{cases} 1 & if \quad (\mathbb{S}^{\mathbf{b}}_s = 1) \wedge (\mathbb{N}^{\mathbf{b}}_s = 1) \wedge \\ & (lifetime \geq T) \\ 0 & otherwise \end{cases}, \tag{10}$$

where $lifetime = l^{\mathbf{b}}_s - f^{\mathbf{b}}_s$ is the amount of frames from the first to the last visualization of $C^{\mathbf{b}}_s$ and $\mathbb{D}^{\mathbf{b}}_s = 1$ denotes stationarity ($\mathbf{b}$ contains a new stable cluster). In case of stationarity, $C^{\mathbf{b}}_s$, $f^{\mathbf{b}}_s$, $l^{\mathbf{b}}_s$ and $w^{\mathbf{b}}_s$ become the last stable representation $\mathcal{S}^{\mathbf{b}}$ and they are included in $\mathcal{O}^{\mathbf{b}}$ as an old visualized cluster.

As stationary objects may spread across several blocks, detections of neighboring locations are associated via connected component analysis. They are visualized as bounding boxes in the result image D_s during a user-defined time.

V. Experimental Results

A. Setup

We use short sequences for evaluation from AVSS07[1] and PETS06[2] datasets as in other works [7][11]. The *PV_M* sequence from AVSS07 is not included as it contains camera jitter and the Online Block Clustering filters every motion block. Moreover, long sequences from CUHK[3], VIRAT[4], IDIAP Traffic Junction[5] and AVSS2007 datasets have been also used for long-term test conditions. CUHK and IDIAP datasets contain one sequence each, while we have merged all the short clips from VIRAT which are continuous in time and useful for SOD, conforming 4 sequences. Overall, 364951 frames (~4.05 h) have been tested and we have visually identified 51 abandoned objects and stopped vehicles as ground-truth[6]. We assume that people are not considered neither false or correct detections as in [8][13] and previous visualization of the empty scene due to the long-term focus of the proposed approach.

[1] http://www.avss2007.org/
[2] http://www.cvg.reading.ac.uk/PETS2006/data.html
[3] http://www.ee.cuhk.edu.hk/~xgwang/CUHK_square.html
[4] http://www.viratdata.org/
[5] http://www.idiap.ch/~odobez/datasets.php
[6] http://www-vpu.eps.uam.es/publications/SOD_STSC/

Table I
COMPARATIVE EVALUATION. GT, TP AND FP DENOTE, RESPECTIVELY, GROUND-TRUTH, CORRECT AND FALSE DETECTIONS. THE PROPOSED APPROACH ACHIEVES BEST RESULTS (BOLD) IN SHORT-TERM AND LONG-TERM SEQUENCES.

Algorithm		Short-term									Long-term									
		AVSS07					PETS06	Mean			AVSS07	IDIAP	VIRAT				CUHK	Mean		
		AB_E	*AB_M*	*AB_H*	*PV_E*	*PV_H*	*Cam3*	P	R	F	*AB_EV*	*I_1*	*V_1*	*V_2*	*V_3*	*V_4*	*C_1*	P	R	F
[6]	GT/TP/FP	1/1/0	1/1/6	1/1/3	1/1/5	1/1/10	1/1/0	.20	1	.33	5/5/21	7/7/10	3/3/8	3/3/10	4/4/10	15/15/6	8/8/20	.35	1	.51
[9]	GT/TP/FP	1/1/0	1/1/0	1/1/0	1/1/0	1/0/2	1/1/0	.71	.83	.77	5/5/5	7/7/2	3/3/6	3/3/6	4/4/9	15/15/0	8/8/10	.54	**1**	.70
[10]	GT/TP/FP	1/1/0	1/1/2	1/1/1	1/1/4	1/1/10	1/1/0	.26	1	.41	5/5/12	7/7/10	3/3/8	3/3/10	4/4/10	15/15/6	8/8/20	.37	1	.54
[11]	GT/TP/FP	1/1/1	1/1/4	1/1/3	1/1/1	1/1/10	1/1/0	.24	1	.39	5/4/25	7/7/6	3/3/0	3/3/0	4/4/0	15/13/3	8/6/7	.49	.89	.63
Proposed	GT/TP/FP	1/1/0	1/1/0	1/1/0	1/1/0	1/1/0	1/1/0	**1**	**1**	**1**	5/5/3	7/6/4	3/3/3	3/3/6	4/4/4	15/14/0	8/8/2	**.66**	.96	**.78**

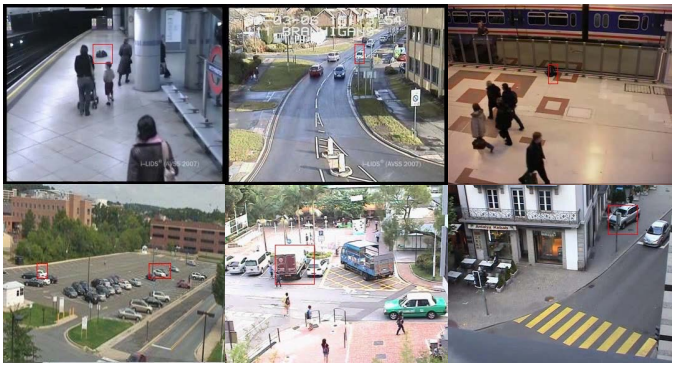

Figure 5. Examples of detections in each dataset. First row, from left to right: *AB_H*, *PV_H* and *C_1*. Second row, from left to right: *I_1*, *Cam3* and *V_4*.

To evaluate the results, we use standard Precision $P = TP/(TP + FP)$, Recall $R = TP/(TP + FN)$ and F-score $F = 2 \cdot P \cdot R/(P + R)$ measures, where TP, FP and FN denote, respectively, correct, false and missed detections.

For the proposed approach, we use N=16 (16×16 blocks) and k=50 frames (sampling rate). The lower (higher) the sampling rate is the lower (higher) the delay in the detections is, thus its value does not have a significant impact on performance. For the online clustering (Section III), we set m=64 as in [16] and z=3 to keep at most 3 clusters for each location $\mathbf{b}$ after each sampling instant. The SVM to match blocks is trained with a balanced set of about 2000 positive and negative samples, i.e. 2000 labeled block comparisons randomly collected from different sequences. Finally, the alarm time T is set to 30 seconds for PETS06 (ground-truth value), 1 minute for AVSS07 (minimum value of the ground-truth ones) and 1 minute for the rest of the datasets without annotations.

B. Comparative evaluation

The proposed approach is compared with four state-of-the-art approaches: temporal accumulation for single [6] and multiple [9] features, temporal sampling [10] and dual background [11]. These approaches are tested both in short-term and long-term sequences. It is fair to mention that other state-of-the-art approaches [7][8] report similar results in the short-term sequences, however as there is no available software to analyze the long-term ones, they have not been included.

The left part of Table I shows the results for short-term sequences. The proposed approach detects all objects without false positives, thus demonstrating the robustness in crowded situations such as *AB_M*, *AB_H* and *PV_H*, where illumination changes and cast shadows take place. The state-of-the-art approaches have serious difficulties to cope with crowds, thus producing false detections in all cases except [9].

The right part of Table I shows the results for long-term sequences which pose additional challenges besides crowds or illumination changes, such as temporal adaptation and correctly handling intermittent object motion. The results show that the proposed approach outperforms the state-of-the-art, specially in sequences which contain intermittent object motion and crowded situations, such as *C_1* and *AB_EV*. This is due to the ability of the Online Block Clustering to perform a fast and illumination-robust adaptation and due to the capability of the Stationary Block Detection stage to identify stationary objects while avoiding the intermittent object motion issue by the buffer $\mathcal{O}^{\mathbf{b}}$. However, the proposed approach fails in few situations, due to uncovered regions of the empty scene where its appearance changes from the one in $\mathcal{O}^{\mathbf{b}}$ and due to camouflage effects. An example of the proposed approach detections is shown in Figure 5.

The computational cost of the proposed approach is mainly due to the clustering stage since the stationary detection performs simple operations each k frames. A non-optimized Matlab implementation runs at 5 *fps* on a standard PC (P-IV 2.8 GHz and 2 GB RAM). Regarding the state-of-the-art [6][9][10][11], the overall cost depends on BS where recent Matlab implementations are in the range 8-10 *fps* [18][19].

VI. CONCLUSIONS

This letter proposes a SOD approach suitable for long-term operation due to its robustness to crowds, illumination changes and intermittent object motion. The proposed approach presents a new strategy to identify stationarity by detecting spatio-temporal stability changes in the scene. The proposed approach employs Online Block Clustering robust to illumination changes, being able to distinguish between equal and different spatial representations of the scene over time. Future work will mainly explore dynamic block size and additional features for the block clustering.

ACKNOWLEDGMENT

This work was partially supported by the Spanish Government (HA-Video TEC2014-5317-R) and by the TEC department (UAM).

REFERENCES

[1] Q. Fan, P. Gabbur, and S. Pankanti, "Relative attributes for large-scale abandoned object detection," in *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, Dec 2013, pp. 2736–2743.

[2] K. Lin, S. Chen, C. Chen, D. Lin, and Y. Hung, "Abandoned object detection via temporal consistency modeling and back-tracing verification for visual surveillance," *IEEE Transactions on Information Forensics and Security*, 2015.

[3] A. Albiol, L. Sanchis, A. Albiol, and J. Mossi, "Detection of parked vehicles using spatiotemporal maps," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1277–1291, Dec 2011.

[4] A. Bayona, J. SanMiguel, and J. Martinez, "Comparative evaluation of stationary foreground object detection algorithms based on background subtraction techniques," in *Proceedings of IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Sept 2009, pp. 25–30.

[5] T. Bouwmans, "Traditional and recent approaches in background modeling for foreground detection: An overview," *Computer Science Review*, vol. 11-12, pp. 31–66, 2014.

[6] S. Guler, J. Silverstein, and I. Pushee, "Stationary objects in multiple object tracking," in *Proceedings of IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Sept 2007, pp. 248–253.

[7] J. Kim and D. Kim, "Accurate static region classification using multiple cues for ARO detection," *IEEE Signal Processing Letters*, vol. 21, no. 8, pp. 937–941, Aug 2014.

[8] J. Pan, Q. Fan, and S. Pankanti, "Robust abandoned object detection using region-level analysis," in *Proceedings of IEEE International Conference on Image Processing (ICIP)*, Sept 2011, pp. 3597–3600.

[9] D. Ortego and J. SanMiguel, "Multi-feature stationary foreground detection for crowded video-surveillance," in *Proceedings of IEEE International Conference on Image Processing (ICIP)*, Oct 2014, pp. 2403–2407.

[10] A. Bayona, J. SanMiguel, and J. Martinez, "Stationary foreground detection using background subtraction and temporal difference in video surveillance," in *Proceedings of IEEE International Conference on Image Processing (ICIP)*, Sept 2010, pp. 4657–4660.

[11] F. Porikli, Y. Ivanov, and T. Haga., "Robust abandoned object detection using dual foregrounds," *EURASIP Journal on Advances in Signal Processing*, vol. Article ID 197875, 2008.

[12] R. Evangelio and T. Sikora, "Complementary background models for the detection of static and moving objects in crowded environments," in *Proceedings of IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Aug 2011, pp. 71–76.

[13] T. YingLi, R. Feris, L. Haowei, A. Hampapur, and S. Ming-Ting, "Robust detection of abandoned and removed objects in complex surveillance videos," *IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews*, vol. 41, no. 5, pp. 565–576, Sept 2011.

[14] G. Szwoch, "Extraction of stable foreground image regions for unattended luggage detection," *Multimedia Tools and Applications*, pp. 1–26, 2014.

[15] Q. Fan, S. Pankanti, and L. Brown, "Long-term object tracking for parked vehicle detection," in *Proceedings of IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Aug 2014, pp. 223–229.

[16] J. Pilet, C. Strecha, and P. Fua, "Making background subtraction robust to sudden illumination changes," in *Proceedings of European Conference on Computer Vision (ECCV)*, vol. 5305, 2008, pp. 567–580.

[17] Q. Wu, H. Cheng, and B. Jeng, "Motion detection via change-point detection for cumulative histograms of ratio images," *Pattern Recognition Letters*, vol. 26, no. 5, pp. 555–563, Apr. 2005.

[18] Z. Chen and T. Ellis, "A self-adaptive gaussian mixture model," *Computer Vision and Image Understanding*, vol. 122, pp. 35–46, 2014.

[19] J.-W. Seo and S. Kim, "Recursive on-line $(2D)^2$PCA and its application to long-term background subtraction," *IEEE Transactions on Multimedia*, vol. 16, no. 8, pp. 2333–2344, Dec 2014.