Detection of Stationary Foreground Objects Using Multiple Nonparametric Background-Foreground Models on a Finite State Machine

Carlos Cuevas, Raquel Martínez, Daniel Berjón, and Narciso García

***Abstract*— There is a huge proliferation of surveillance systems that require strategies for detecting different kinds of stationary foreground objects (e.g., unattended packages or illegally parked vehicles). As these strategies must be able to detect foreground objects remaining static in crowd scenarios, regardless of how long they have not been moving, several algorithms for detecting different kinds of such foreground objects have been developed over the last decades. This paper presents an efficient and high-quality strategy to detect stationary foreground objects, which is able to detect not only completely static objects but also partially static ones. Three parallel nonparametric detectors with different absorption rates are used to detect currently moving foreground objects, short-term stationary foreground objects, and long-term stationary foreground objects. The results of the detectors are fed into a novel finite state machine that classifies the pixels among background, moving foreground objects, stationary foreground objects, occluded stationary foreground objects, and uncovered background. Results show that the proposed detection strategy is not only able to achieve high quality in several challenging situations but it also improves upon previous strategies.**

***Index Terms*— Stationary foreground object, abandoned object, removed object, background subtraction, nonparametric modeling, background, foreground, finite state machine.**

I. Introduction

DETECTING stationary foreground objects (i.e. foreground objects that become static) is a key task in many video surveillance systems for public security. Some typical scenarios are, for example, the detection of unattended packages in a railway station or in an airport [1], [2], the detection of stolen objects in a museum [3], the identification of abandoned objects on roads [4], or the detection of illegally parked vehicles [5]. Moreover, the detection of stationary foreground objects allows improving the quality of foreground object detection strategies in scenarios featuring objects that stop moving frequently (e.g. people in offices or vehicles on urban roads). In these kinds of scenarios, typical detection strategies lead to frequent misdetections that can be avoided by detecting not only the moving foreground objects but also the foreground objects that temporarily remain static [6], [7].

A. Contribution

A high-quality strategy for detecting all kinds of stationary foreground objects is proposed. First, three independent background-foreground nonparametric modeling-based detectors with different absorption rates are applied on each input image. The first one detects only the moving foreground objects. The second one also detects the foreground objects that have recently stopped. The third one detects the moving foreground objects and the stationary foreground objects (regardless of how long they have not been moving). Finally, the results provided by the detectors are used as input of an efficient Finite State Machine (FSM) that classifies the pixels among background, moving foreground object, stationary foreground object, uncovered background and occluded stationary foreground object.

One of the most important contributions in this paper is the use of nonparametric kernel density estimation (KDE) detectors. The usual methods for detecting stationary foreground (e.g. parametric methods) summarize the history of each pixel. However, in constrast, KDE-based methods explicitly store the most recent values of each pixel, which in turn allows to factor in the temporal distance from each reference datum to the input, thus enabling the detection of only those foreground objects that remain static for a certain time. In this paper, the proposed KDE-based models include an innovative selective update mechanism to control the absorption rate of each detector. In addition, whereas the learning rates of the models in parametric strategies must be adapted by the users to the characteristics of the analyzed sequence, the proposed selective update allows using a single configuration whatever the content of the sequences. Therefore, the proposed strategy is more useful than strategies based on parametric methods.

Another important contribution of the proposed strategy is the simultaneous use of three KDE-based models with different learning rates, which allows to deal with complex situations (e.g. occluded foreground) without requiring a complex FSM. To be able to detect these complex situations, previous works including an FSM only use two moving object detectors [8]–[10], mainly due to the complex configuration required by the parametric models they use. However,

Manuscript received November 23, 2015; revised June 28, 2016 and September 12, 2016; accepted December 14, 2016. Date of publication December 20, 2016; date of current version January 20, 2017. This work was supported by the Ministerio de Economía, Industria y Competitividad of the Spanish Government under Project TEC2013-48453 (MR-UHDTV) and Project TEC2016-75981 (IVME). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Aleksandra Pizurica.

The authors are with the Grupo de Tratamiento de Imágenes, Information Processing and Telecommunications Center (IPTC) and ETSI Telecomunicacioìn, Universidad Politécnica de Madrid, 28040 Madrid, Spain (e-mail: ccr@gti.ssr.upm.es; rms@gti.ssr.upm.es; dbd@gti.ssr.upm.es; narciso@gti.ssr.upm.es).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2016.2642779

1057-7149 © 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

using only two detectors requires complex FSMs to provide successful classifications. Since the proposed selectively updated KDE-based models can be configured much more easily, using more than two independent detectors becomes practical. Moreover, the addition of a third model simplifies the design of the FSM. Nevertheless, although it has been found that the addition of a third model simplifies the design of the FSM, using more than three detectors increases the computational cost without achieving better classification results.

II. Related Work

Over the last two decades, a significant amount of works describing strategies to detect stationary foreground objects have been proposed [11].

Most of these strategies are focused on detecting abandoned objects [12], since the detection of this kind of objects is a key task in security applications for preventing terrorist incidents and to reduce crime [13]. In many cases, the proposed strategies just detect abandoned objects [14], [15]. However, some works include mechanisms to also identify the owners of the abandoned objects [16]–[18]. Thus, they can determine, depending on whether the owner stays with the object or not, if an abandonment can result in a dangerous situation.

There are also approaches focused on detecting stopped vehicles, since this is of great interest for traffic monitoring applications and parking surveillance [19]. Some approaches analyze the behavior of vehicles on roads [20], others focus on urban scenarios [5], while others consider mixed scenarios (both roads and cities) [21].

Some works are not limited to the detection of a particular type of object, but try to detect any moving foreground object that becomes static [22]–[24]. Additionally, some of these works not only consider the detection of abandoned objects or stopped vehicles, but also the detection of people remaining totally or partially static [25]–[27].

A significant share of the strategies in the literature is based on tracking algorithms that try to determine which foreground regions stop moving [19]. The simplest ones try to relate objects across pairs of consecutive frames by analyzing colors, distances, velocities, or object sizes [15]. Other strategies are based on higher-level information [28] or use standard tracking algorithms (e.g. Kalman-based tracking in [29], a pyramidal Kanade-Lucas-Tomasi algorithm in [30], or particle filters in [31]).

Although tracking-based strategies are able to provide successful detections in many scenarios, they typically require establishing the characteristics of the objects to track and, additionally, because they work at object-level, they are not able to detect partially static foreground objects (i.e. people that only move the upper body). Consequently, over the past few years, most authors have opted for pixel-level strategies, which are based on an initial foreground segmentation by a foreground object detector. Some of these pixel-based strategies are based on persistence analyses, whereas others are based on dual foreground comparisons.

Among the strategies based on persistence analyses, the simplest ones conclude that a pixel is part of a stationary foreground object when it is classified as foreground for a predetermined lapse of time [32], [33] or along several frames (consecutive [14], [26] or not [23], [34]). Other strategies, instead of directly analyzing the persistence of the result provided by the foreground detector, analyze the stability of the Gaussians associated to each pixel in a Gaussian Mixture Model (GMM) [35]. When a foreground object appears in a pixel, a new Gaussian is created in its GMM representing the new value of the pixel. If the object stops moving, this new Gaussian begins to gain importance in the mixture. So, by identifying this situation, it is possible to determine when a foreground object becomes static. This idea was first proposed in [25] and, virtually simultaneously, also in [36] (with small differences between them). Later, it has been incorporated into other strategies [37]–[39] that are able to detect both totally and partially static foreground objects. Nevertheless, they are unable to detect long-term stationary foreground objects. Additionally, they fail in complex scenarios with dynamic backgrounds.

The approaches based on dual foreground comparisons take as starting point the strategy published in [40], which proposes constructing two binary foreground masks from two background models with different learning rates. The model with the fastest learning rate (commonly called short-term model) must be configured to adapt rapidly to the changes in the scene, so it only detects short duration changes (i.e. the objects in motion). In contrast, the model with the slowest learning rate (long-term model) must be configured to be more resistant against changes. So, it must also detect long duration changes (i.e. the stationary foreground objects). In the strategy proposed in [40] and some other later works [41]–[43], the two models are constructed using GMMs. However, other modeling choices can also be found in the literature: nonstatistical models in [44]–[46], single Gaussian models in [47], median models in [48], or cluster models in [49]. Many of these strategies are commonly able to provide successful detections in scenarios with complex backgrounds and detect partially and totally static foreground objects. However, they are not able to maintain the detections when the objects remain static for a long time and they lose the detected stationary objects when other foreground objects pass in front of them. Additionally, to provide successful results, the configurations of the long-term and short-term models must be adapted to the characteristics of each analyzed sequence. Thus, the usability of these methods is low.

All the previously described detection strategies are based on foreground masks obtained from a background modeling stage. If the background models are updated with a blind update mechanism, the objects remaining stationary for long periods of time cannot be detected because, sooner or later, these objects are always absorbed by the background models. On the other hand, if a selective update is used, the long-term stationary foreground objects can be correctly detected. However, a selective update mechanism does not allow to distinguish between stationary and moving foreground objects. To detect long-term stationary foreground objects, as well to distinguish them from the moving ones, many authors use traditional detection methods in conjunction with an FSM. For example, in [50], a strategy that uses dual foreground com-

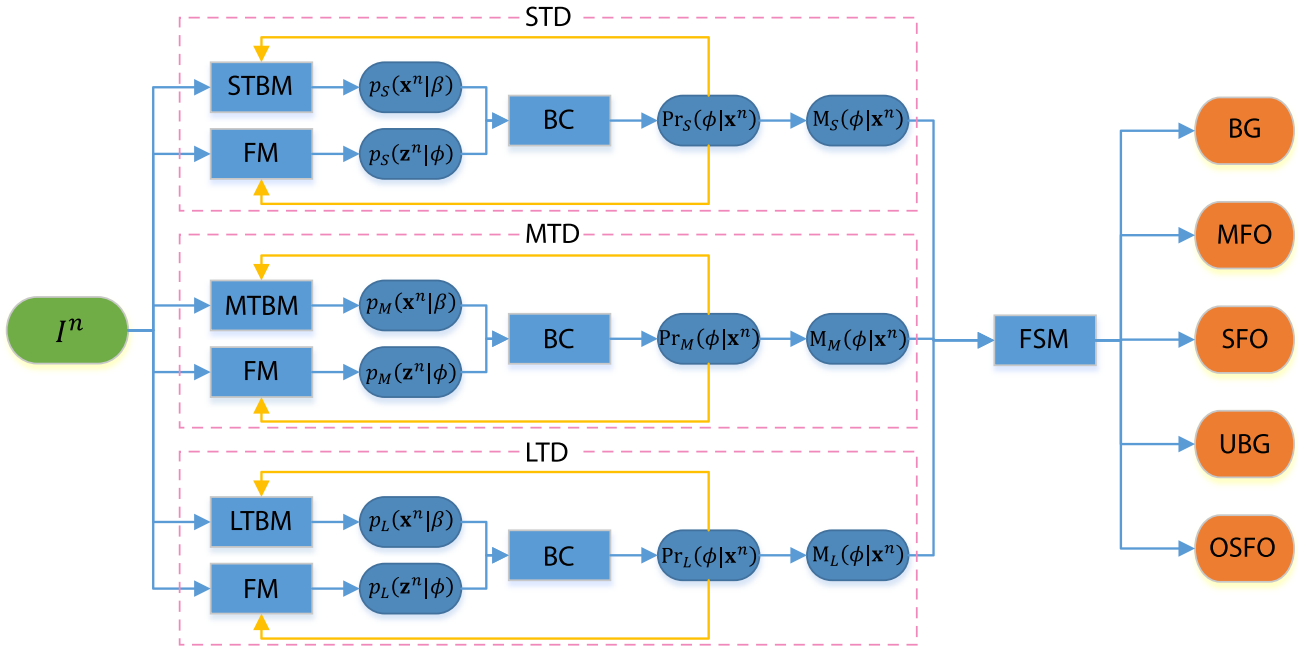

Fig. 1. Block diagram of the proposed system. Notation: round-edged blocks denote data and rectangular blocks denote processes. The green and orange blocks indicate, respectively, the input and the output of the system.

parisons and an FSM was proposed. An extended description of this strategy was later proposed in [8]. In [51], an FSM is supplied with the results provided by a tracking module. In [52], the FSM is used in conjunction with the results of a persistence analysis (improved versions of this work were later published in [9] and [53]). The persistence analysis in [54] is also supported by an FSM. Finally, in [10] an FSM is combined with a dual foreground comparison to detect candidate stationary foreground objects and a tracker is then used to verify whether such candidates are really abandoned objects or not.

III. System Overview

The proposed strategy, depicted in Fig. 1, comprises two main stages. First, a robust foreground detection using KDE-based nonparametric background and foreground models is performed. Then, an efficient FSM is used to determine which foreground objects remain static.

For each new frame, I^n, at time n, the foreground objects in the scene are detected by using three motion detectors with different absorption rates. The first detector, called "short-term detector" (STD), detects only those foreground objects that are in motion in the current frame. The second one, called "medium-term detector" (MTD), also detects those foreground objects that have recently stopped. Finally, the third detector, called "long-term detector" (LTD), detects the foreground objects in motion and all the foreground objects remaining static (regardless of how long they have not been moving).

Three nonparametric background models and three spatio-temporal nonparametric foreground models are used to perform these detections. The background models differ in the way in which they are updated: the first one, called "short-term background model" (STBM), makes use of a selective update mechanism that rapidly absorbs those foreground objects that stop moving; the selective update used in the second model, called "medium-term background model" (MTBM), results in a slower absorption of such stationary foreground objects; finally, thanks to the selective update used in the third model, called "long-term background model" (LTBM), the stationary foreground objects are never absorbed by the model. Each of the three background models is combined with the corresponding foreground model (FM) in a Bayesian classifier (BC) to obtain a probability of each image pixel belonging to the foreground of the sequence. Note that the absorption rates are the only difference between the proposed background models, and that there are no differences between the foreground models (all their parameters are configured with identical values).

The probabilities resulting from the detectors are thresholded to obtain three binary masks (M_S, M_M and M_L). These masks are inputs of an efficient FSM that classifies each pixel into five classes: background (BG), moving foreground object (MFO), stationary foreground object (SFO), uncovered background (UBG) and occluded stationary foreground object (OSFO).

IV. Foreground Detection

Each of the three proposed detectors is based on comparing a spatio-temporal nonparametric foreground model [55] and a nonparametric background model that includes an efficient selective update mechanism to easily configure the absorption rate desired for each modeling. In contrast to previous methods using multiple foreground detectors, the proposed selective update allows using the same configuration whatever the content of the analyzed sequence.

A. Background Modeling

Let $\mathbf{x}^n$ be a D-dimensional vector containing the appearance information of a pixel, p^n, in the current image, I^n, at time n. Let $\{\mathbf{x}_\beta^i\}_{i=1}^{N_\beta}$ be a set of N_β reference samples obtained from the pixels at the same coordinates of p^n in the T_β previous images. Applying Gaussian kernels with diagonal covariance matrices, $\Sigma_{\beta,\mathbf{x}^n} = \text{diag}(\sigma_{\beta,1}^2, \sigma_{\beta,2}^2 \ldots \sigma_{\beta,D}^2)$, the pdf that p^n belongs to the image background, β, is estimated as

$$\text{p}(\mathbf{x}^n|\beta) = \frac{1}{\sum_{i=1}^{N_\beta} w_i} \cdot \frac{1}{(2\pi)^{\frac{D}{2}}} \cdot \frac{1}{\left|\Sigma_{\beta,\mathbf{x}^n}\right|^{\frac{1}{2}}} \cdot \sum_{i=1}^{N_\beta} w_i \prod_{j=1}^{D} \exp\left(-\frac{\left(\mathbf{x}^n(j) - \mathbf{x}_\beta^i(j)\right)^2}{2\Sigma_{\beta,\mathbf{x}^n}(j,j)}\right), \quad (1)$$

where w_i is a weight, assigned to the i-th reference sample, determined by the selective update mechanism that is detailed in section V.

The kernel widths are dynamically estimated as proposed in [56],

$$\Sigma_{\beta,\mathbf{x}^n}(j,j) = \frac{m_j}{0.68\sqrt{2}} : j \in [1, D], \quad (2)$$

where m_j is the median of the absolute values of the differences between the $j-$th component of the consecutive reference samples, $\left\{\left|\mathbf{x}_\beta^i(j) - \mathbf{x}_\beta^{i-1}(j)\right|\right\}_{i=2}^{N_\beta}$.

B. Foreground Modeling

Presuming that foreground objects are commonly in motion, the foreground pdf that a pixel p^n belongs to the image foreground should be estimated not only from previous pixels at the same spatial position of p^n but also from previous pixels at different coordinates. Therefore, to compute the foreground modeling, both the vector defining the current pixel and the foreground reference samples must take into account the spatial coordinates of the data [55].

Let $\mathbf{z}^n = ((\mathbf{x}^n)^T, (\mathbf{s}^n)^T)^T$ be a $D+2$-dimensional vector, where $\mathbf{x}^n$ is the appearance vector described in subsection IV-A and $\mathbf{s}^n = (h^n, w^n)$ is a vector containing the spatial coordinates (row and column) of p^n. Let $\left\{\mathbf{z}_\phi^i\right\}_{i=1}^{N_\phi}$ be the set of N_ϕ reference samples classified as foreground in the T_ϕ previous images into a spatial neighborhood around (h^n, w^n). The pdf that p^n belongs to the image foreground, ϕ, is estimated as

$$\text{p}\left(\mathbf{z}^n|\phi\right) = \alpha\gamma + \frac{1-\alpha}{N_\phi(2\pi)^{1+\frac{D}{2}}} \cdot \frac{1}{\left|\Sigma_{\phi,\mathbf{z}^n}\right|^{\frac{1}{2}}} \cdot \sum_{i=1}^{N_\phi} \prod_{j=1}^{D+2} \exp\left(-\frac{\left(\mathbf{z}^n(j) - \mathbf{z}_\phi^i(j)\right)^2}{2\Sigma_{\phi,\mathbf{z}^n}(j,j)}\right), \quad (3)$$

where $\alpha \ll 1$ is a mixture factor, γ is a constant density of a uniform random variable in the $D+2$ components defined for the feature vector $\mathbf{z}^n$, and $\Sigma_{\phi,\mathbf{x}^n} = \text{diag}\left(\sigma_{\phi,1}^2, \sigma_{\phi,2}^2 \ldots \sigma_{\phi,D}^2, \sigma_{\phi,H}^2, \sigma_{\phi,W}^2\right)$ is the covariance matrix used in the kernels.

The spatial width values used in this modeling depend on the number of reference images, T_ϕ, and on the speed of foreground objects, i.e. they should be large enough to take into account, for each foreground object, all the relevant reference data in all the reference images [57]. Details on the values assigned to these widths are provided in section VII.

Since the distribution of reference data is not dense, the widths of the appearance components cannot be determined using the same procedure of the background. Therefore, these widths must be manually set. If the widths are too large, the objects with similar appearance to the background of the scene will not be correctly detected. On the other hand, if the selected widths are smaller than the widths used in the background, false detections will feed back and become persistent. The values assigned to these widths are also detailed in section VII.

C. Bayesian Classifier

On the one hand, the background models are obtained using only appearance information. However, on the other hand, the foreground modeling is computed by using appearance and spatial data. Therefore, instead of the typical Bayesian classifier [58], an alternative one that allows decoupling the appearance and spatial information is used:

$$\text{Pr}\left(\phi|\mathbf{x}^n\right) = \frac{\text{p}\left(\mathbf{x}^n|\phi, \mathbf{s}^n\right)}{\text{p}\left(\mathbf{x}^n|\phi, \mathbf{s}^n\right) + \text{p}\left(\mathbf{x}^n|\beta\right)} \quad (4)$$

where $\text{p}\left(\mathbf{x}^n|\phi, \mathbf{s}^n\right)$ results from conditioning the foreground model, $\text{p}\left(\mathbf{z}^n|\phi\right)$, on a particular spatial location. This conditioned density function is obtained as

$$\text{p}\left(\mathbf{x}^n|\phi, \mathbf{s}^n\right) = \frac{\text{p}\left(\mathbf{z}^n|\phi\right)}{\text{p}\left(\mathbf{s}^n|\phi\right)} \quad (5)$$

where $\text{p}\left(\mathbf{s}^n|\phi\right)$ is the marginalization of $\text{p}\left(\mathbf{z}^n|\phi\right)$ over the D-dimensional set of appearance characteristics. This marginal density function is obtained as

$$\text{p}\left(\mathbf{s}^n|\phi\right) = \alpha\gamma' + \frac{1-\alpha}{N_\phi 2\pi} \cdot \frac{1}{\sigma_{\phi,H}\sigma_{\phi,W}} \cdot \sum_{i=1}^{N_\phi} \prod_{j=D+1}^{D+2} \exp\left(-\frac{\left(\mathbf{z}^n(j) - \mathbf{z}_\phi^i(j)\right)^2}{2\Sigma_{\phi,\mathbf{z}^n}(j,j)}\right), \quad (6)$$

where γ' is a constant density in the spatial components.

D. Foreground Masks

Let $\text{Pr}_\text{S}\left(\phi|\mathbf{x}^n\right)$, $\text{Pr}_\text{M}\left(\phi|\mathbf{x}^n\right)$ and $\text{Pr}_\text{L}\left(\phi|\mathbf{x}^n\right)$ denote the probabilities of p^n of being part of the foreground, provided by the short-term, medium-term and long-term detectors. These probabilities are thresholded as

$$M_\zeta\left(\phi|\mathbf{x}^n\right) = \begin{cases} 1 & \text{if } \text{Pr}_\zeta\left(\phi|\mathbf{x}^n\right) \geq 0.5 \\ 0 & \text{if } \text{Pr}_\zeta\left(\phi|\mathbf{x}^n\right) < 0.5 \end{cases} : \zeta \in \{S, M, L\} \quad (7)$$

to obtain binary data indicating if each pixel is classified as part of the foreground ($M_\zeta\left(\phi|\mathbf{x}^n\right) = 1$) or part of the background ($M_\zeta\left(\phi|\mathbf{x}^n\right) = 0$). These binary data will be the input of the FSM described in section VI.

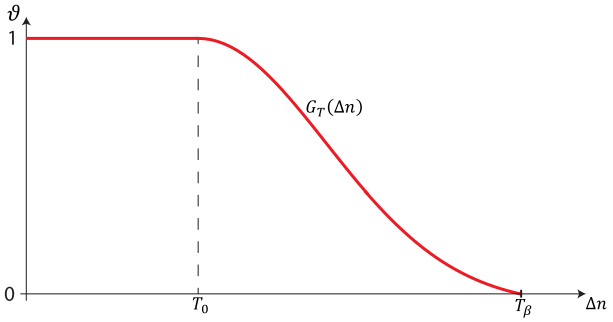

Fig. 2. Temporal weight used to perform the selective update of the background models.

V. Selective Update

The only difference between the three background models is the way in which their selective update is performed. This update is controlled by the weights mentioned in subsection IV-A, which are obtained as

$$w_i = 1 - \vartheta \Pr\left(\phi|\mathbf{x}_\beta^i\right), \tag{8}$$

where $\Pr\left(\phi|\mathbf{x}_\beta^i\right)$ is the probability assigned to the reference sample $\mathbf{x}_\beta^i$ of being part of the foreground and ϑ is a temporal weight obtained as

$$\vartheta = \begin{cases} 1 & \text{if } \Delta n \leq T_0 \\ G_T\left(\Delta n\right) & \text{if } \Delta n > T_0 \end{cases}. \tag{9}$$

In this temporal weight, illustrated in Fig. 2, Δn is the temporal distance between the reference samples and the current one, $T_0 < T_\beta$ is a predefined constant value that controls the absorption rate of the model and $G_T\left(\Delta n\right)$ is a temporal Gaussian defined as

$$G_T\left(\Delta n\right) = \exp\left(-\frac{(\Delta n - T_0)^2}{2\sigma_T^2}\right). \tag{10}$$

To guarantee that the value of this Gaussian in T_β is approximately 0 (i.e. $G_T\left(T_\beta\right) \approx 0$), its standard deviation must be set as

$$\sigma_T \leq \frac{T_\beta - T_0}{3} \tag{11}$$

Thus, on the one hand, the contribution of the T_0 most recent background reference samples only depends on their probability to be classified as background, $\Pr\left(\beta|\mathbf{x}_\beta^i\right) = 1 - \Pr\left(\phi|\mathbf{x}_\beta^i\right)$. Therefore, assuming that, at a given time, the foreground objects are correctly detected, the samples of such objects will just barely affect the background model corresponding to the following T_0 images. Consequently, even if the foreground objects stop moving, during such period their absorption by the background model will be negligible (the absorption could vary slightly depending on the exact probability values assigned to the reference samples).

On the other hand, the probability values associated with the reference samples with $\Delta n > T_0$ lose relevance in the modeling as the value of Δn increases. Therefore, all the reference samples (whether they have been classified as foreground or background) will end up influencing the model. In this way, foreground objects remaining static will gradually become part of the background. In other words, the proposed scheme treats recent reference samples as a selective update and distant past samples as a blind update, with a soft transition between them.

To work correctly, the proposed strategy requires that the following two conditions are satisfied simultaneously:

- The STD must absorb the stationary foreground objects noticeably faster than the MTD.
- The LTD should never absorb the stationary foreground objects.

Let $T_{0,S}$, $T_{0,M}$ and $T_{0,L}$ denote the values of T_0 assigned to, respectively, the STD, MTD and LTD. The second condition is easily satisfied if $T_{0,L} = T_\beta$ (pure selective update), whereas the first condition is satisfied if $T_{0,S} < T_{0,M} < T_\beta$.

The value of $T_{0,S}$ must be higher than the number of frames that a pixel is covered by the moving objects in the scene. Therefore, it depends on the size and speed of the moving objects and on the frame-rate of the sequence (the higher the frame-rate, the bigger the moving objects and/or the lower their speed, the higher this value must be set). Regarding $T_{0,M}$, it must be considered that if the difference between $T_{0,S}$ and $T_{0,M}$ is too small, slow-moving foreground objects could be erroneously classified as stationary foreground objects. Conversely, if such difference is too large or if the value of $T_{0,M}$ is too high, only those foreground objects remaining static very long periods of time will be detected (i.e. short-term stationary foreground objects will be misdetected). Typically, using $T_{0,M} = 2T_{0,S}$ is a good compromise between the fast detection of SFOs and avoid false classifications due to excessively slow moving objects.

The results obtained with the proposed detectors in two different scenarios and using the same configuration (i.e. the same values of T_0) are illustrated in Fig. 3. The first scenario shows a person that has stopped in front of a door whose color is similar to that of his sweatshirt. The second sequence shows the typical abandonment of a backpack (the displayed image shows a portion of the person who has left the backpack). It can be observed that the STD has significantly absorbed the two stationary foreground objects, whereas the MTD has absorbed them much less. In addition, the figure shows that the LTD has not absorbed any of the foreground objects.

VI. Finite State Machine

In the second stage of the proposed system, the binary data obtained as described in subsection IV-D are introduced in an FSM. This machine will classify each current pixel among five classes: background, moving foreground object, stationary foreground object, uncovered background (removed object), and stationary foreground object occluded by another foreground object.

All the FSMs included in other SMO detection strategies take as input two binary masks resulting from two foreground detection algorithms with different update speeds. However, the proposed FSM takes as input three detection masks. Thus, in contrast to previous FSMs, the proposed one is able to obtain successful classifications by using very few states, which significantly simplifies the operation of the machine.

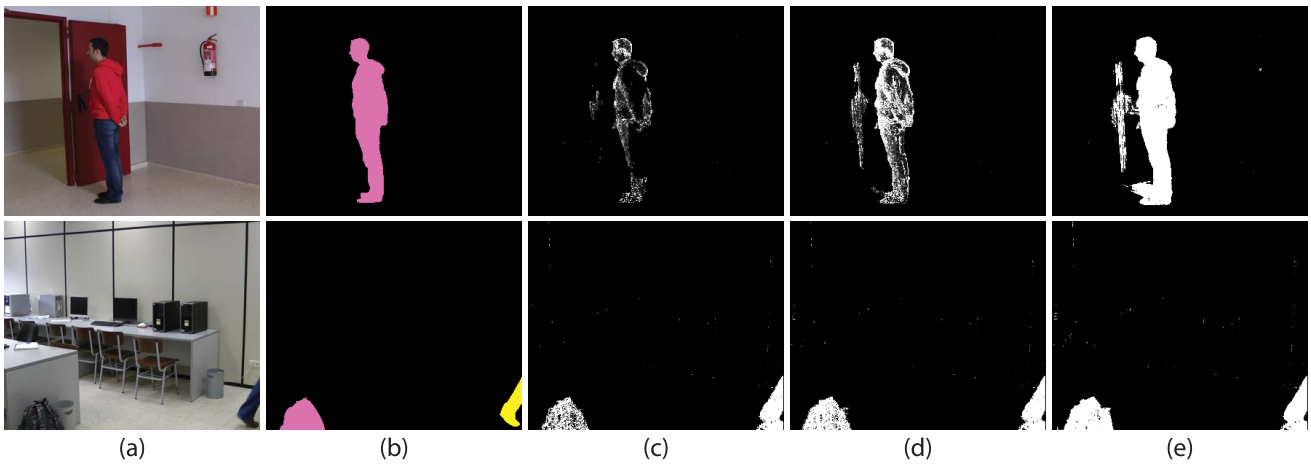

Fig. 3. Results obtained with the proposed nonparametric detectors. (a) Original images. (b) Ground-truth. (c) Results from the STD ($T_{0,S} = 20$). (d) Results from the MTD ($T_{0,M} = 40$). (e) Results from the LTD ($T_{0,L} = T_\beta$). Color notation for the ground-truth images (this notation will be the same for the rest of figures containing ground-truth images): The moving foreground objects are depicted in yellow and the stationary foreground objects are depicted in pink.

Moreover, whereas other FSMs require the use of auxiliary variables to avoid premature changes between states, the proposed FSM does not depend on any parameter. Therefore, its usability is very high.

An FSM can be defined as a 5-tuple $(S, Q, Z, \delta, \varpi)$ [59], where:

- S is the input alphabet.
- Q is the set of states in the machine.
- Z is the output alphabet.
- δ is a function that, depending on the current state and the current input, determines the next state.
- ϖ is the output function that, depending on the current state and the current input, determines the output of the machine.

In the case of the proposed FSM, the elements of this 5-tuple are defined as follows:

- S is the set of possible combinations of the 3-tuple $(M_\mathrm{L}, M_\mathrm{M}, M_S)$.
- Q is the set of 5 states described in subsection VI-A.
- Z is a number (from 0 to 4) indicating the pixel classification: 0 for background pixels, 1 for pixels belonging to moving foreground objects, 2 for pixels belonging to stationary foreground objects, 3 for uncovered background and 4 for pixels belonging to occluded stationary foreground objects.
- δ is the next-state function illustrated in Fig. 4.
- ϖ is a function with output values $z \in \{0, 1 \ldots 4\}$ corresponding to the state of a pixel at a given time.

A. State Description

The proposed FSM is composed by 5 states numbered from 0 to 4. Their descriptions and the conditions that are necessary to reach them are:

- State 0 (BG - Background): This is the initial state for every pixel and it denotes that the pixel is part of the background of the scene. It is reached when a pixel is classified as background by all three detectors

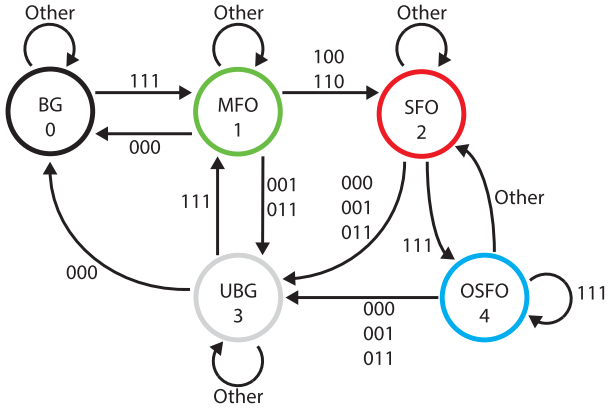

Fig. 4. Next-step function of the proposed FSM.

(i.e. $(M_\mathrm{L}, M_\mathrm{M}, M_S) = (0, 0, 0)$) and the previous state was MFO or UBG.

- State 1 (MFO - Moving foreground object): The pixels in this state are classified as part of a moving foreground object. It is reached when the pixel is classified as foreground by all three detectors (i.e. $(M_\mathrm{L}, M_\mathrm{M}, M_S) = (1, 1, 1)$) and the previous state was BG or UBG.
- State 2 (SFO - Stationary foreground object): It denotes that the pixels are part of a stationary foreground object. It is reached when a foreground object remains static along several consecutive frames and, consequently, it is absorbed by the STD or the MTD, but not by the LTD (i.e. $(M_\mathrm{L}, M_\mathrm{M}, M_S) = (1, 0, 0)$ or $(M_\mathrm{L}, M_\mathrm{M}, M_S) = (1, 1, 0)$). This state is also reached when a pixel in the state OSFO (stationary foreground objects that are covered by other foreground objects) is uncovered (i.e. $(M_\mathrm{L}, M_\mathrm{M}, M_\mathrm{S}) \neq (1, 1, 1)$, with $M_L = 1$).
- State 3 (UBG - Uncovered background): The pixels in this state are classified as part of an uncovered region (i.e. a stationary foreground object moves again or a background object is removed by someone). To reach this state, a pixel in the states MFO, SFO or OSFO must be classified as foreground by one of the detectors, while

being classified as background by a longer-term detector (i.e. $(M_{\mathrm{L}}, M_{\mathrm{M}}, M_S) = (0, 1, 1)$ or $(M_{\mathrm{L}}, M_{\mathrm{M}}, M_{\mathrm{S}}) = (0, 0, 1)$).

- State 4 (OSFO - Occluded stationary foreground object): The pixels classified in this state belong to stationary foreground objects that are temporally occluded by other foreground objects. It is reached when a pixel is classified as a stationary foreground object and then, the three detectors classify it as part of the foreground (i.e. $(M_{\mathrm{L}}, M_{\mathrm{M}}, M_S) = (1, 1, 1)$).

VII. Results

To assess the quality and robustness of the proposed strategy, a large set of sequences, which contains many typical challenges in stationary foreground detection (long-term abandonments, foreground objects that remain partially static, occluded stationary foreground objects, etc.), has been used. These sequences have been extracted from the following four databases:

- PETS2006[1] [60]: It is the most widely used database in the literature to assess the quality of strategies for detecting abandoned objects. It was designed to test the detection of abandoned luggage in seven scenarios of different complexity. The scenarios were filmed from multiple cameras. However, since the proposed strategy does not consider the use of multiple cameras, similarly to many other previous works [39], [46], [50], only the sequences captured with the frontal camera (view 3) have been used in the performed experiments. These sequences, labeled S1 to S7, contain different kinds of abandoned objects and also foreground objects that remain partially static. In addition, in many of them the abandoned objects are temporally occluded by other foreground objects.
- i-LIDS[2] [61]: This database is the second most used in the literature to test strategies for detecting stationary foreground. It contains two sets of sequences to test, respectively, the detection of abandoned baggage (AB sequences) and illegally parked vehicles (PV sequences). Each set is conformed by three sequences of different difficulty (easy, medium and high). All the sequences are supplied with XML files describing temporal events (alarms). In the case of the first set, the alarms start sixty seconds after the person that has placed the baggage on the floor leaves the vicinity of such baggage. The alarms end when the baggage is recovered by his owner. In the second set, an alarm starts sixty seconds after a vehicle remains stationary on a no parking zone and the alarm stops when the vehicle moves again. Since the proposed strategy does not include any high-level stage to establish relations between foreground objects (i.e. between the abandoned objects and their owners), only the second set of this database (PV sequences) has been considered in the performed experiments.
- LASIESTA[3] [62]: This database stands out among others because it is the only existing database with real videos that are fully annotated at both pixel and object levels. Additionally, it is the only one including a specific label for stationary foreground objects. The sequences in LASIESTA are distributed into many categories addressing different challenges in moving object detection.
- ChangeDetection[4] [63]: This database has reached a great popularity since its emergence in 2012. It contains 49 video sequences classified in 10 categories related to typical challenges in moving object detection. Along with LASIESTA, it is the only one providing ground-truth data at pixel level.

It must be noted that PETS2006 and i-LIDS were specifically created to evaluate the performance of stationary foreground object detectors. LASIESTA and ChangeDetection, on the other hand, are generic databases that include sequences for addressing not only the challenge of detecting stationary foreground but many more challenges (robustness against shadows, illumination changes, dynamic background, moving cameras, etc.). However, these two databases are the only ones providing pixel-level foreground masks, which allows a quantitative analysis of the quality of the results provided by the evaluated detection strategies. Moreover, such masks can be used not only to provide measures related to the detection of SFOs but to evaluate the quality in the detection of generic foreground objects (i.e. stationary or in motion). Since it is the focus of this paper, only those sequences containing stationary foreground objects have been considered. In the case of the LASIESTA database, there are three sequences with this kind of objects. Two of them (named "I_CA_01" and "I_CA_02") contain foreground objects remaining partially static. The third one (named "I_MB_01") shows a typical baggage abandonment. In the case of the ChangeDetection dataset, there is a category named "Intermittent Object Motion" that was specifically created for evaluating the quality of the detectors when the foreground is not always in motion. This category is composed by six sequences. However, four of them are bootstrapping sequences (they contain objects in the initial background that are removed by someone throughout the sequence). The proposed algorithm has not been designed to detect such removal events. Consequently, only the two remaining sequences in the mentioned category have been finally used. These sequences are named "Sofa" and "StreetLight".

The evaluation of the quality of the proposed detection strategy has been accomplished through three experiments. The first one (subsection VII-C) is focused on the analysis of the speed in detecting foreground objects that have stopped moving, as well as in the ability of the strategy for maintaining the detection of both long-term and occluded stationary foreground objects. The second experiment (subsection VII-D) aims to assess the quality of the strategy in the detection of not only stationary foreground but also moving foreground. Finally, the third experiment (subsection VII-E) has been

[1] http://www.cvg.reading.ac.uk/PETS2006
[2] http://www.eecs.qmul.ac.uk/~andrea/avss2007_d.html
[3] http://www.gti.ssr.upm.es/data/LASIESTA
[4] http://www.changedetection.net

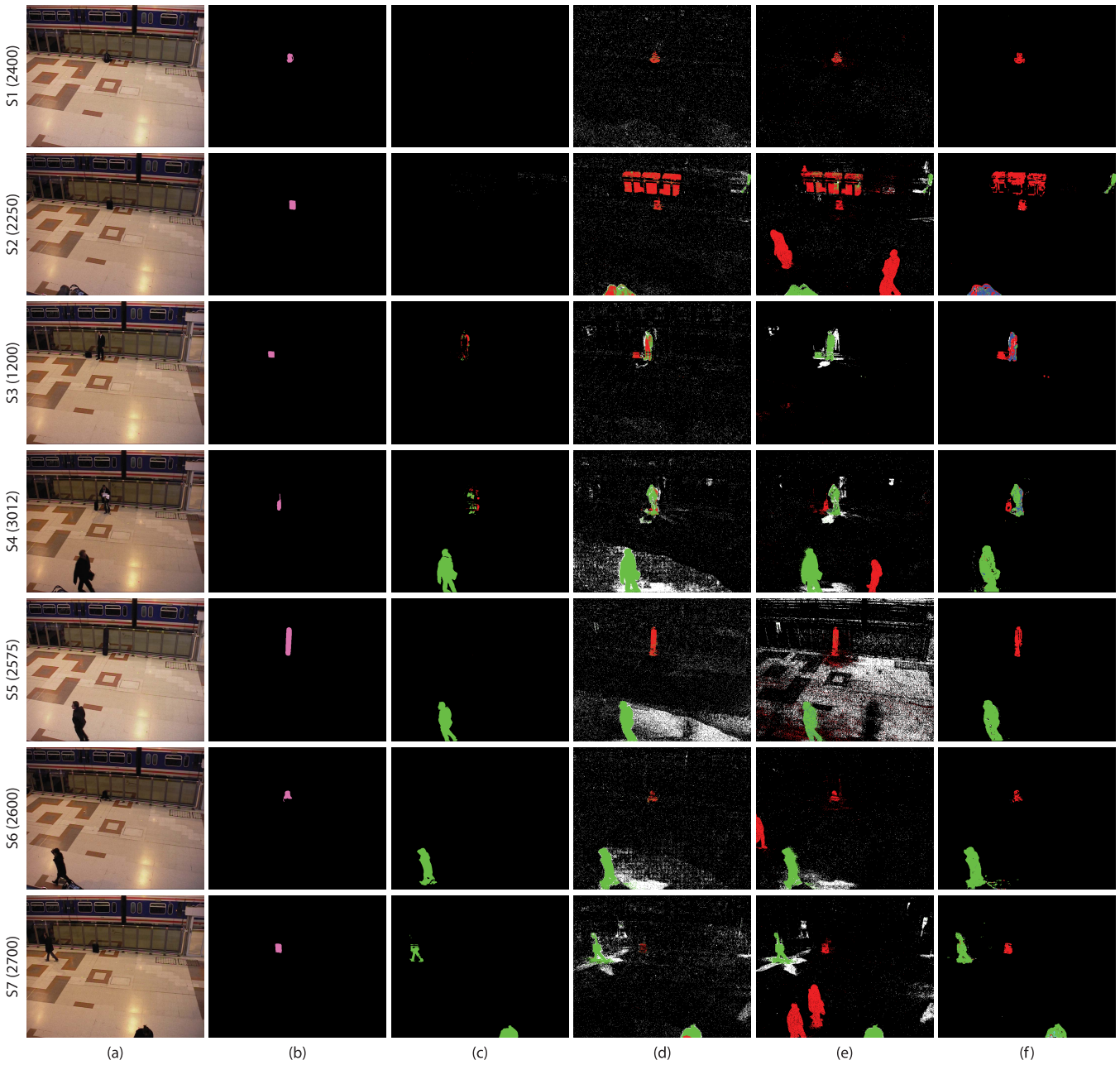

Fig. 5. Some representative results obtained on PETS2006 dataset. (a) Original images. (b) Ground-truth. (c) Results provided by the strategy GS. (d) Results obtained with the strategy DFC. (e) Results obtained with the strategy DFC-FSM. (f) Results obtained with the proposed strategy. Color notation in the results: SFO in red, MFO in green, UBG in white, OSFO in blue and BG in black.

designed to evaluate the ability of the proposed strategy to generate alarms that must be triggered when a foreground object stops for longer than a specified duration.

All test sequences, their ground-truth and the obtained results are available at a public website.[5]

A. Parameter Selection

To reduce the influence of shadows and reflected light in the detections, all the nonparametric models are obtained using the appearance vector described in [64], which is composed by the chromaticity (Rn, Gn) and the module of the gradient of the brightness, $|\nabla S|$.

[5]http://www.gti.ssr.upm.es/data/

In the case of the background models, only values of T_β and T_0 must be established. The first one has been set to $T_\beta = 600$, which is more than enough to model the cyclical background changes in all the test sequences. Regarding the values of T_0, the performed experiments have shown that using $T_{0,S} = 20$ and $T_{0,M} = 40$ all the requirements discussed in section V are satisfied and, as it is shown in the following subsections, successful results are obtained in all the evaluated scenarios.

In the case of the foreground models, it is necessary to set the number of reference images, the spatial width of the kernels and the appearance width of the kernels. The former has been set as $T_\phi = 10$, which is typically enough in any sequence, since the foreground does not typically exhibit

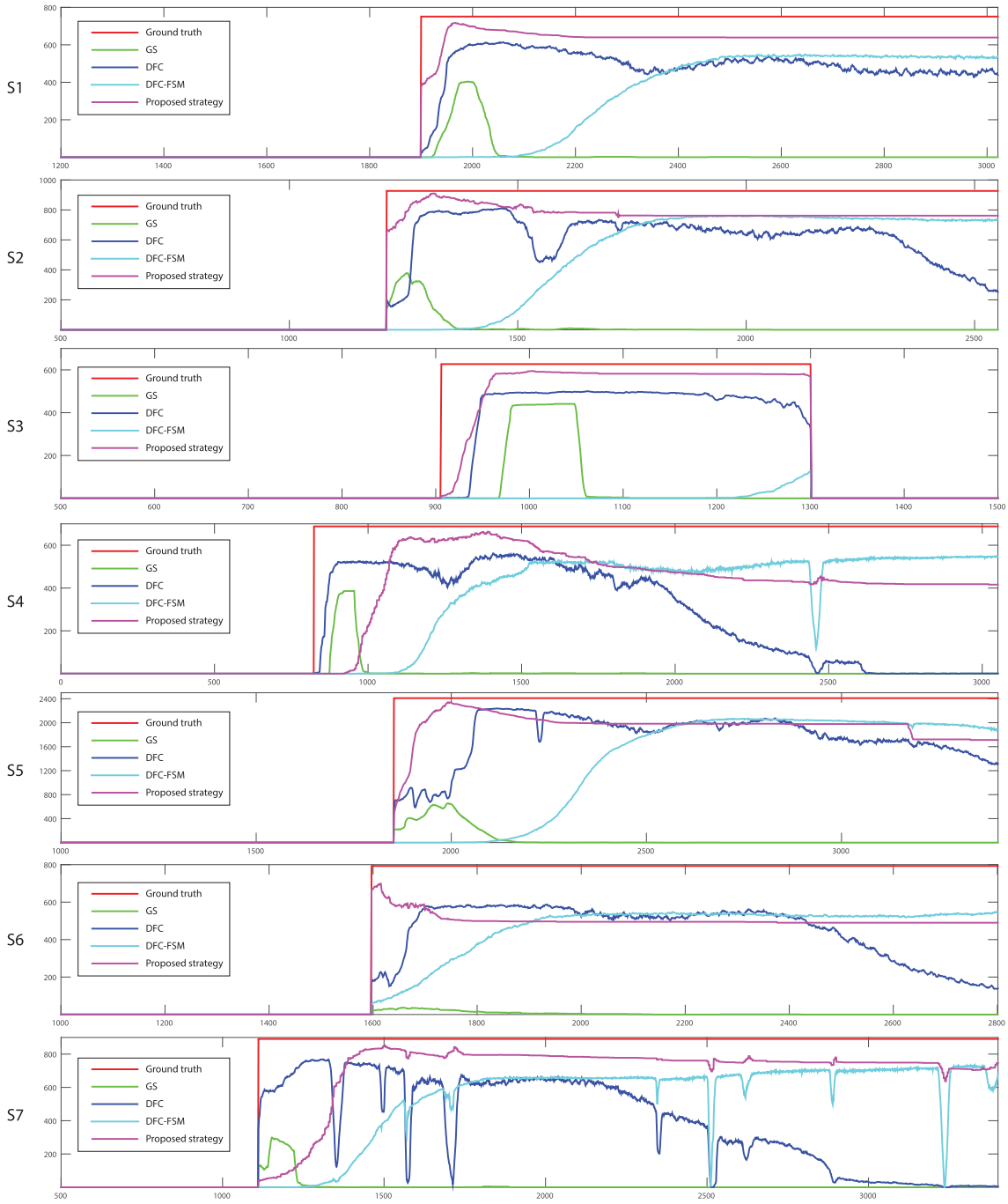

Fig. 6. Amount of pixels correctly classified as SFO along the sequences of PETS2006 dataset. These curves allow to see the detection speed of the strategies that are compared and also their ability to maintain the detections when the objects of interest are occluded.

cyclical changes as the background may do. The spatial widths of the kernels have been set as $\sigma^2_{\phi,H} = \sigma^2_{\phi,W} = \left(\frac{5}{3}\right)^2$. Finally, the appearance width has been set as 0.02, which is sufficiently larger than the typical values for the background noise, which are in the order of 10^{-3} in the set of appearance components that has been used.

B. Computational Analysis

The proposed strategy has been implemented on a general-purpose graphics processing unit (GPGPU) nVidia GTX-580.

TABLE I

MEAN COMPUTATIONAL COST PER FRAME IN EACH STAGE OF THE PROPOSED STRATEGY (288 × 352 RESOLUTION)

Stage	BG modeling	FG modeling	Overall
Cost	13 ms	14 ms	27 ms

Table I shows the computational cost of each stage. The background and foreground models have been implemented taking as starting point the implementation described in [57]. In the case of the background models, since their only difference

lies in the use of different weights (see eq. (1)), most of the computational load is shared among them. Therefore, the use of three models instead of one does not result in a triple cost, but much less. Regarding the foreground, the input data of each model depend on previous results. Therefore, in this case it is not possible to save cost. That is, the cost associated to the three foreground models triples the cost of a single model. Finally, it must be noted that the computational cost related to the FSM is negligible compared to the costs in the previous stages of the strategy.

C. Experiment 1: Speed and Robustness Against Occlusions

This experiment is focused on measuring the detection speed of the proposed strategy and proving its ability to maintain the detection of stationary foreground objects when they are occluded by other foreground objects. To this end, the results obtained with the proposed strategy have been compared with the results provided by the two most used pixel-level strategies for detecting SFOs. The first strategy (henceforth "Gaussian Stability (GS)") is based on an analysis of the stability of the Gaussians in a GMM associated to each pixel. This strategy was first proposed in [36] and, because of its computational efficiency and its high-quality results in many complex scenarios, it has been used by several authors over the past few years [37], [39]. The second strategy (henceforth "Dual Foreground Comparison (DFC)") is based on the dual foreground comparison proposed in [40], which has been taken as starting point by many recent works [43], [46]. In contrast to the proposed strategy, which can be successfully applied on all the test sequences by using a single set of parameters (see subsection VII-A), both of these detection methods are highly dependent on the selection of an adequate learning rate for the GMMs. Taking this into account, these methods have been configured with the best learning rate that has been found for each sequence (although this puts the proposed strategy at a disadvantage compared to these algorithms).

Additionally, the obtained results have also been compared with those provided by the strategy in [50] (henceforth "DFC-FSM"), which takes as starting point the dual foreground in [40] but, similarly to the proposed strategy, includes an FSM to try to improve the quality of the detections in sequences with objects that remain static for a long time and sequences where the stationary foreground objects are occluded by other foreground objects.

Many of the above referenced works, after applying the pixel-level detection, include different region-level post-processing steps that could also be applied to the proposed strategy. In this experiment, only the quality of the initial pixel-level detection stage is analyzed. Therefore, the use of these post-processing steps has not been considered.

The databases selected to carry out these comparisons have been PETS2006 and LASIESTA. On the one hand, the seven sequences in PETS2006 show long-term abandoned objects that, in many cases, are occluded by other foreground objects. On the other hand, the sequences in LASIESTA contain foreground objects that stop moving for a while and then resume their motion.

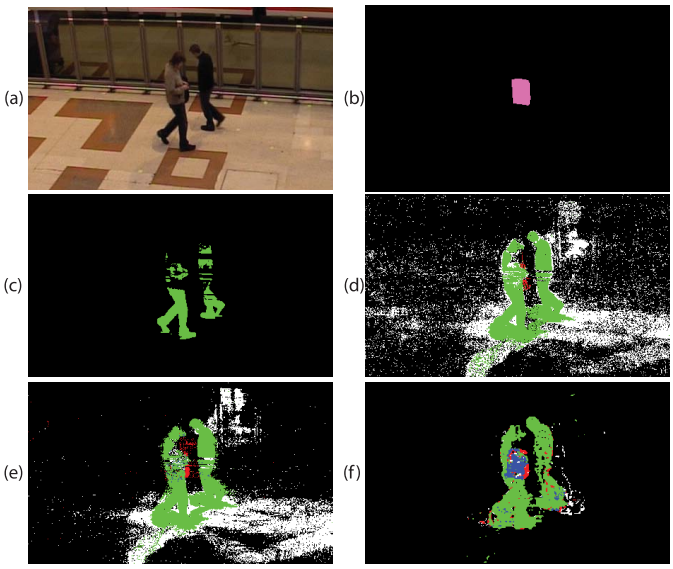

Fig. 7. Abandoned bag occluded by other foreground objects. (a) Original image. (b) Ground-truth. (c) Results provided by the strategy GS. (d) Results provided by the strategy DFC. (e) Results provided by the strategy DFC-FSM. (f) Results provided by the proposed strategy. Color notation in the results: SFO in red, MFO in green, UBG in white, OSFO in blue and BG in black.

1) Performance on the PETS2006 Dataset: Some representative results obtained with the proposed strategy and the aforementioned alternative strategies on the sequences of PETS2006 are illustrated in Fig. 5. The amount of pixels correctly classified as SFO along such sequences are also illustrated in the graphics of Fig. 6.

As it can be seen in these figures, the proposed strategy usually obtains the highest amounts of pixels correctly classified as SFO. On the one hand, the strategies GS and DFC are not able to correctly classify those objects remaining static very long periods of time (this problem is especially severe in the case of the strategy GS). On the other hand, thanks to the use of FSMs, the strategy DFC-FSM and the proposed one are able to maintain the detections of the SFOs regardless of how long they remain static. However, the strategy DFC-FSM takes much longer to detect the presence of a SFO.

Moreover, as can be observed in the graphics corresponding to sequences S4 and S7, the proposed strategy is the only one that is able to maintain the detection of SFOs when they are occluded. The rest of the strategies exhibit significant reductions in the amount of pixels correctly classified as SFO. The images in Fig. 7 illustrate the results obtained by the four strategies under study when an abandoned bag is occluded by another foreground object. It can be observed that the proposed strategy is the only one that is able to maintain the correct classification of such abandoned bag.

Finally, as most of the examples in Fig. 5 show, the proposed strategy is not only able to provide the highest amounts of correct detections but also avoids erroneous classifications: it can be observed that strategies DFC and DFC-FSM erroneously classify as UBG significant amounts of pixels. These erroneous classifications are mainly due to the fact that the sequences contain continuous lighting changes due to many reasons (e.g. camera auto-adjustments or people in

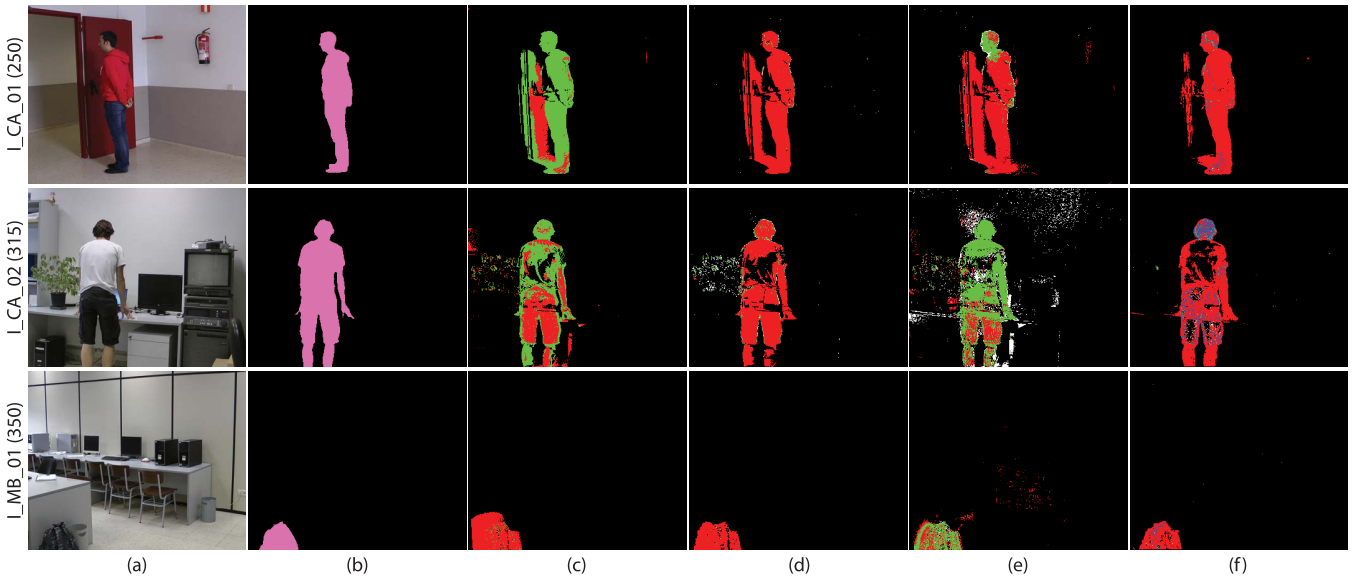

Fig. 8. Some representative results obtained on the LASIESTA database. (a) Original images. (b) Ground-truth. (c) Results provided by the strategy GS. (d) Results obtained with the strategy DFC. (e) Results provided with the strategy DFC-FSM. (f) Results obtained with the proposed strategy. Color notation in the results: SFO in red, MFO in green, UBG in white, OSFO in blue and BG in black.

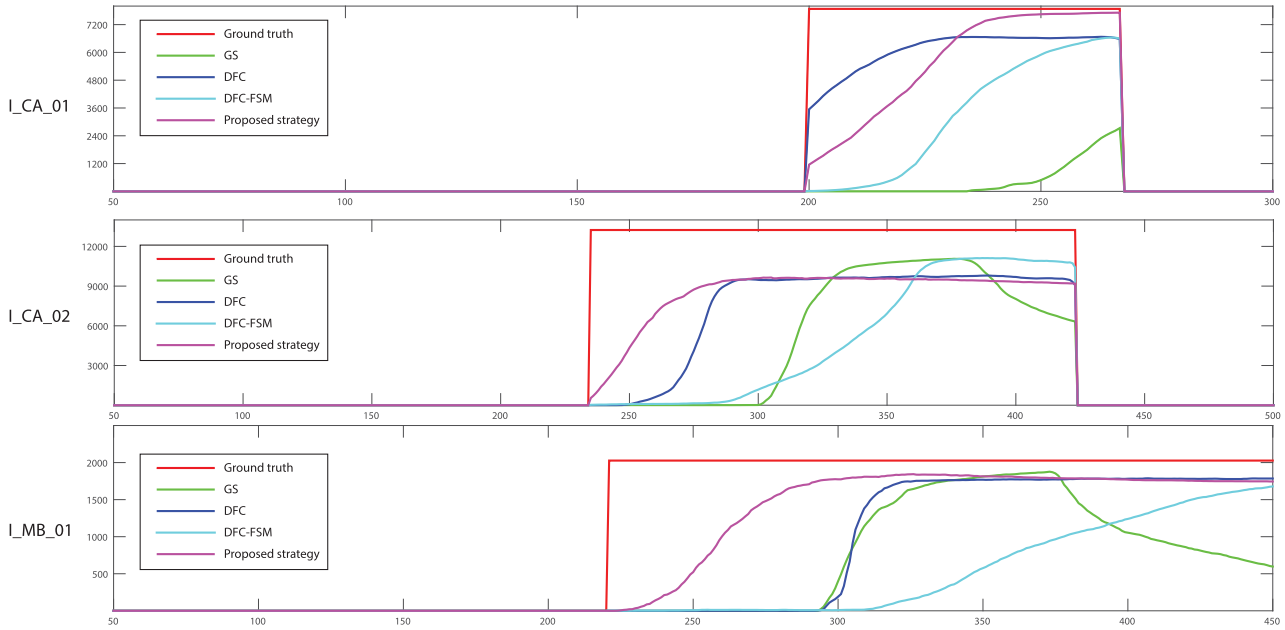

Fig. 9. Amount of pixels correctly classified as SFO along the sequences of the LASIESTA database. These curves show the speed of the strategies that are compared to detect the SFOs.

the scene causing shadows). These changes are frequently detected as foreground by GMMs with high learning rates (short-term models) but not by GMMs with low learning rates (long-term models). Consequently, the strategies that are based on comparing short and long-term GMMs erroneously classify as UBG those pixels that have been affected by the lighting changes. Nevertheless, it must be noted that the pixels erroneously classified as UBG because of lighting changes are irrelevant in the evaluation of the quality of the algorithms, since they are treated as what they ultimately are, that is, background.

2) Performance on the LASIESTA Database: Some representative results obtained with the proposed strategy and the aforementioned alternative strategies on the three sequences selected from LASIESTA are illustrated in Fig. 8. The amounts of pixels correctly classified as SFO along such sequences are illustrated in Fig. 9. In these sequences, the proposed strategy is the one that detects the SFOs best (it classifies correctly more pixels than the rest of the strategies and it also detects the SFOs faster). Again, strategies GS and DFC are not able to maintain the detection of SFOs when the objects remain static for too many consecutive frames, whereas the

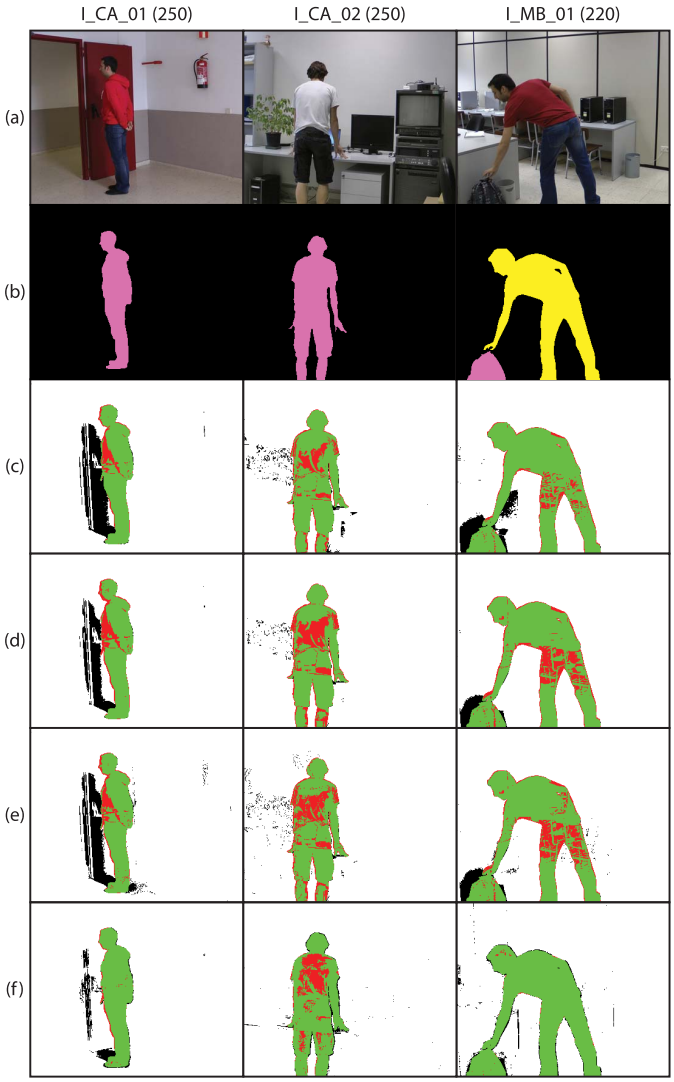

Fig. 10. Some representative results in the LASIESTA database considering the joint detection of MFOs and SFOs. (a) Original images. (b) Ground-truth. (c) Results with the strategy GS. (d) Results with the strategy DFC. (e) Results with the strategy DFC-FSM. (f) Results with the proposed strategy. Color notation in the results: TP in green, FP in black and FN in red.

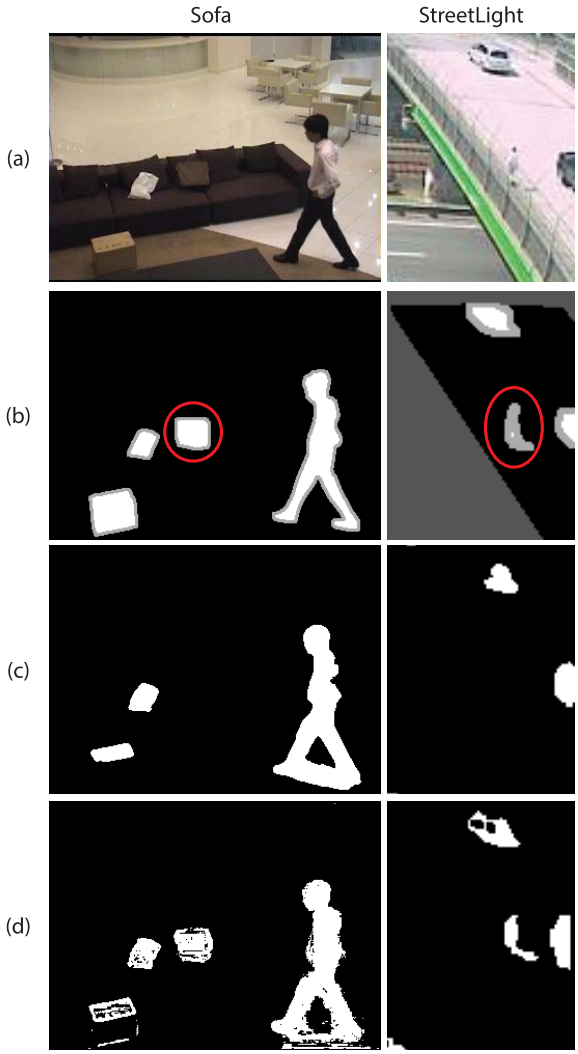

Fig. 11. Some representative results in the ChangeDetection database considering the joint detection of MSOs and SFOs. (a) Original images. (b) Ground-truth (as provided by the ChangeDetection dataset). (c) Results with the strategy FTSG. (d) Results with the proposed strategy.

strategy DFC-FSM takes much longer for detecting a SFO (the examples illustrated in Fig. 8 show that many pixels of SFOs remain erroneously classified as part of MFOs).

D. Experiment 2: Detection of MFOs and SFOs

The aim of this experiment is to evaluate the quality of the proposed strategy in the detection, at pixel level, of both moving and stationary foreground objects. To carry out this evaluation, the LASIESTA and the ChangeDetection databases have been used, since they are the only ones providing pixel-level labels for both types of foreground. LASIESTA is a very recent database and, to our knowledge, it has not been previously used to assess the quality of any strategy for stationary foreground detection. Hence, in this database we have compared the results obtained with the proposed strategy to those achieved with the three strategies described in the previous experiment (GS, DFC and DFC-FSM). On the other hand, ChangeDetection is a more consolidated database and it has been previously used to assess the quality of only one strategy (henceforth "FTSG" [27]) for detecting stationary foreground. Therefore, the results of the proposed strategy in this dataset have been compared to those obtained with such strategy. It must be highlighted that, currently, this strategy is located at the first position in the ranking of the best strategies assessed with the ChangeDetection database.

This experiment has been done by using the conventional recall (r) and precision (p) evaluation parameters,

$$r = 100\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}\%, \qquad p = 100\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}\%, \tag{12}$$

where TP (true positive) is the amount of pixels correctly classified as foreground (i.e. as MFO, SFO or OSFO), FN (false negative) is the number of foreground pixels that have not been classified as foreground, and FP (false positive) is the amount of pixels erroneously classified as foreground. Additionally, their harmonic mean or F-score ($F = 2\frac{r \cdot p}{r+p}$) has been used to jointly evaluate the recall and the precision.

TABLE II

START AND END TIMES (IN MINUTES) OF THE ALARMS OBTAINED IN I-LIDS WITH THE PROPOSED STRATEGY AND OTHER APPROACHES

	PV-Easy		PV-Medium		PV-Hard		Mean	Median
	Start	End	Start	End	Start	End	error	error
GT	02:48	03:15	01:28	01:47	02:12	02:33	-	-
2007-Boragno [65]	02:48	03:19	01:28	01:55	02:12	02:36	5.0	4
2007-Guler [66]	02:46	03:18	01:28	01:54	02:13	02:36	5.3	5
2007-Venetianer [67]	02:52	03:16	01:43	01:47	02:19	02:34	9.3	8
2008-Porikli [68]	n/a	n/a	01:39	01:47	n/a	n/a	11.0	11
2009-Lee [19]	02:51	03:18	01:33	01:52	02:16	02:34	7.0	6
2013-Maddalena [26]	02:45	03:19	01:28	01:51	02:12	02:34	4.0	4
Proposed	02:48	03:17	01:29	01:51	02:12	02:34	2.7	2

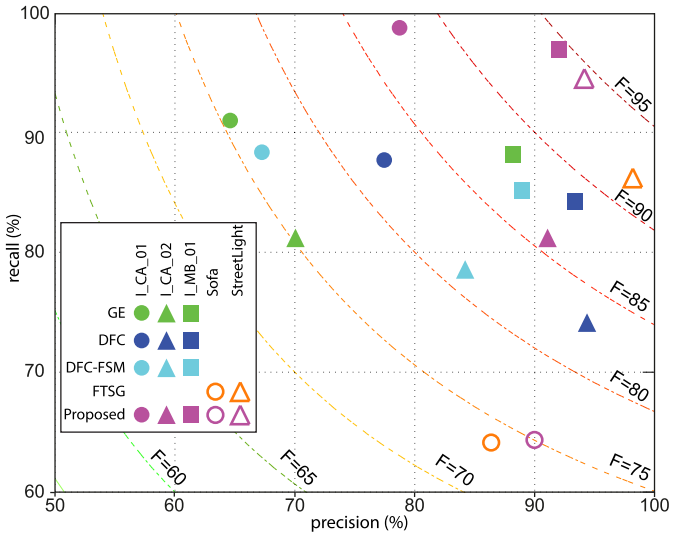

Fig. 12. Quantitative evaluation (SFOs+MFOs) between the proposed strategy and the alternative ones in LASIESTA (solid geometric shapes) and ChangeDetection (hollow geometric shapes). The curve lines represent some isopercentages of the F-score.

Fig. 13. Some representative results obtained in i-LIDS. (a) Original images. (b) Pixel-level results obtained with the proposed strategy (SFO in red, MFO in green, UBG in white, OSFO in blue and BG in black). (c) Object-level masks after applying morphological filtering. (d) Persistence masks (number of consecutive frames during which each pixel has been classified as SFO or OSFO). The events that must be detected are highlighted with pink ellipses. The blue ellipse shows a SFO that does not cause an alarm.

Some representative images obtained in this experiment are illustrated in Fig. 10 (for the results in LASIESTA) and Fig. 11 (for the results in ChangeDetection). Additionally, the obtained recall-precision percentages and their corresponding F-scores are shown in Fig. 12. In the case of the results obtained in LASIESTA, it can be observed that the proposed nonparametric background-foreground modeling provides the highest amounts of correct detections (the highest recall percentages). In addition, it also avoids many false detections due to dynamic changes in the background (for example, the GMM-based algorithms are not able to model the changes in the plant in the background of the second example illustrated in Fig. 10). Consequently, the F-scores obtained with the proposed strategy are significantly higher than those achieved by the rest of evaluated strategies. Regarding the results obtained in the ChangeDetection database, the proposed strategy is also able to beat the quality of the strategy FTSG, better discriminating between background and foreground, even if the foreground objects are very small or are camouflaged (see the objects marked by the red ellipses in Fig. 11).

E. Experiment 3: Alarm Handling

This experiment is focused on demonstrating that the proposed strategy is capable of generating temporal alarms with very high precision. For that purpose the i-LIDS database has been used, since it has been specifically designed to evaluate the accuracy in the generation of alarms at specific times after the abandonment of objects. In addition, there are many other approaches in the literature that have also used this database to evaluate the temporal precision of their results.

In contrast to the previous experiments, in this one it is necessary to analyze the obtained results at object level. Consequently, some typical post-processing operations (morphological opening and closing filters) have been applied to group pixels in blobs. Additionally, it has been necessary to define persistence masks in which the values of the pixels denote the number of consecutive frames during which each

TABLE III

AMOUNT OF TRUE POSITIVES (TP) AND FALSE POSITIVES (FP), AT OBJECT-LEVEL, OBTAINED IN I-LIDS WITH THE PROPOSED STRATEGY AND OTHER APPROACHES

	PV-Easy		PV-Medium		PV-Hard	
	TP	FP	TP	FP	TP	FP
2011-Albiol [5]	1	n/a	1	n/a	1	n/a
2011-Pan[69]	1	0	1	0	1	0
2011-Tian [38]	1	0	1	0	1	1
2013-Maddalena [26]	1	0	1	0	1	0
2015-Filonenco [46]	1	1	1	0	1	0
Proposed	1	0	1	0	1	0

pixel has been classified as part of a SFO. An alarm is triggered if the persistence mask contains values exceeding a predefined threshold. The i-LIDS dataset establishes that an alarm must start 60 seconds after a vehicle stops. Consequently, the threshold has been set to 1500 (the sequences have been recorded at 25 fps).

Table II summarizes the start and end times of the alarms generated by the proposed strategy and many other previous approaches for detecting SFOs. The data in this table show that the proposed strategy achieves the best temporal precision in the generation of the alarms. Thanks to the proposed selective update and to the ability of the KDE-based modeling for taking into account only the recent history of the pixels, the SFOs are detected very fast. Moreover, they are also reclassified as MFOs very shortly after they start moving.

The images in Fig. 13 illustrate some representative results obtained with the proposed strategy. The first one (left column) shows a frame in which there is a vehicle that stopped more than 60 seconds ago (highlighted with a pink ellipse) and a second vehicle that stopped at a crossroads about 20 seconds ago. Although both cars are identified as SFOs, the persistence mask allows to determine that only the first one must result in an alarm. The images in the second example (middle column) correspond to a moment in which a vehicle just stopped. Although it has stopped very recently, it can be observed that it has already been correctly classified as a SFO. However, its value in the persistence mask is too low to generate an alarm. Finally, the last example (right column) illustrates the results obtained when a stopped vehicle that generated an alarm has just resumed its motion. It can be noted that in this case the alarm is about to disappear (i.e. there are almost no pixels with value higher that 1500 in the persistence mask), which proves the speed of the proposed strategy also for stopping alarms.

Finally, Table III shows the obtained results in terms of events correctly detected (true positives) and undetected events (false positives). These results show that the proposed strategy provides a 100% of both recall and precision. There are other previous approaches (e.g. Maddalena [26]) that also provide the same quality. However, they typically require to use a specific set of parameters in each sequence. Conversely, the results obtained with the proposed strategy have been achieved using the same set of parameters in all the sequences, which proves its great usability compared to other previous strategies.

VIII. CONCLUSIONS

A high-quality strategy for detecting stationary foreground objects at pixel level has been proposed. This strategy is suitable for detecting foreground objects that are totally or partially static in a large variety of complex situations (e.g. long-term stationary foreground objects, dynamic backgrounds, camouflage, or stationary foreground objects occluded by other foreground objects).

First, three background-foreground nonparametric detectors with different absorption rates allow detecting, respectively, moving foreground objects, short-term stationary foreground objects, and long-term stationary foreground objects. The absorption rates of these detectors are easily controlled thanks to an efficient selective update mechanism that allows using the same configuration whatever the content of the analyzed sequence, thus offering better usability than previous strategies also based on detectors with different learning rates.

Then, the outputs provided by the three detectors are used as input of a simple and efficient finite state machine that classifies the pixels among background, moving foreground objects, stationary foreground objects, uncovered background and occluded stationary foreground objects. This machine allows to correctly classify the stationary foreground objects regardless of how long they have remained stationary and, additionally, it is able to maintain the detections when the stationary objects are occluded by other foreground objects.

The proposed strategy has been tested on a wide variety of sequences containing critical situations. The obtained results have shown that the proposed strategy is able to provide successful classifications in many challenging scenarios and that it significantly improves upon the results provided by previous strategies for detecting stationary foreground objects.

Despite the successful results obtained in many sequences from four databases, there are some issues that the proposed strategy is not able to deal with: bootstrapping sequences (sequences starting with background objects that are removed by someone) and multi-layered stationary objects (i.e., stationary objects covering other stationary objects). In the future it is intended to complete the proposed strategy to be able to deal with these problems. In the case of bootstrapping sequences, it would be possible to improve the quality of the results by adding additional object-level stages to discriminate between removals and abandonments. Regarding the case of multi-layered stationary objects, it could probably be addressed by including "parallel" finite state machines that would start working once a pixel were classified as part of an occluded stationary object.

REFERENCES

[1] S. Lu, J. Zhang, and D. Feng, "A knowledge-based approach for detecting unattended packages in surveillance video," in *Proc. IEEE Int. Conf. Video Signal Based Surveill.*, Nov. 2006, p. 110.

[2] S.-N. Lim and L. S. Davis, "A one-threshold algorithm for detecting abandoned packages under severe occlusions using a single camera," Dept. Comput. Sci. Eng., Univ. Maryland, College Park, College Park, MD, USA, Tech. Rep.CS-TR-4784, 2006.

[3] S. Ferrando, G. Gera, and C. Regazzoni, "Classification of unattended and stolen objects in video-surveillance system," in *Proc. IEEE Int. Conf. Video Signal Based Surveill.*, Nov. 2006, p. 21.

[4] Y. Zeng, J. Lan, B. Ran, J. Gao, and J. Zou, "A novel abandoned object detection system based on three-dimensional image information," *Sensors*, vol. 15, no. 3, pp. 6885–6904, 2015.

[5] A. Albiol, L. Sanchis, A. Albiol, and J. M. Mossi, "Detection of parked vehicles using spatiotemporal maps," *IEEE Trans. Intell. Transp. Syst.*, no. 4, pp. 1277–1291, Dec. 2011.

[6] T. Bouwmans, "Traditional and recent approaches in background modeling for foreground detection: An overview," *Comput. Sci. Rev.*, vols. 11–12, pp. 31–66, May 2014.

[7] A. Sobral and A. Vacavant, "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos," *Comput. Vis. Image Understand.*, vol. 122, pp. 4–21, May 2014.

[8] R. H. Evangelio and T. Sikora, "Complementary background models for the detection of static and moving objects in crowded environments," in *Proc. IEEE Int. Conf. Adv. Video Signal-Based Surveill.*, Aug./Sep. 2011, pp. 71–76.

[9] Q. Fan, P. Gabbur, and S. Pankanti, "Relative attributes for large-scale abandoned object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2736–2743.

[10] K. Lin, S.-C. Chen, C.-S. Chen, D.-T. Lin, and Y.-P. Hung, "Abandoned object detection via temporal consistency modeling and back-tracing verification for visual surveillance," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 7, pp. 1359–1370, Jul. 2015.

[11] C. Cuevas, R. Martínez, and N. García, "Detection of stationary foreground objects: A survey," *Comput. Vis. Image Understand.*, vol. 152, pp. 41–57, Nov. 2016.

[12] K. Sajith and K. N. R. Nair, "Abandoned or removed objects detection from surveillence video using codebook," *Int. J. Eng. Res. Technol.*, vol. 2, no. 5, pp. 1–7, May 2013.

[13] J. Kim and B. Kang, "Nonparametric state machine with multiple features for abnormal object classification," in *Proc. IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Aug. 2014, pp. 199–203.

[14] R. K. Tripathi, A. S. Jalal, and C. Bhatnagar, "A framework for abandoned object detection from video surveillance," in *Proc. Nat. Conf. Comput. Vis., Pattern Recognit., Image Process. Graph.*, Dec. 2013, pp. 1–4.

[15] F. S. Mahin, M. N. Islam, G. Schaefer, and M. A. R. Ahad, "A simple approach for abandoned object detection," in *Proc. Int. Conf. Image Process., Comput. Vis., Pattern Recognit.*, 2015, p. 427.

[16] G. Dalley, X. Wang, and W. E. L. Grimson, "Event detection using an attention-based tracker," in *Proc. Int. Workshop Perform. Eval. Tracking Surveill.*, 2007, pp. 71–79.

[17] L. Chang, H. Zhao, S. Zhai, Y. Ma, and H. Liu, "Robust abandoned object detection and analysis based on online learning," in *Proc. IEEE Int. Conf. Robot. Biomimetics*, Dec. 2013, pp. 940–945.

[18] J. Ferryman *et al.*, "Robust abandoned object detection integrating wide area visual surveillance and social context," *Pattern Recognit. Lett.*, vol. 34, no. 7, pp. 789–798, May 2013.

[19] J. T. Lee, M. S. Ryoo, M. Riley, and J. K. Aggarwal, "Real-time illegal parking detection in outdoor environments using 1-D transformation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 7, pp. 1014–1024, Jul. 2009.

[20] R. Melli, A. Prati, R. Cucchiara, and L. de Cock, "Predictive and probabilistic tracking to detect stopped vehicles," in *Proc. WACV/MOTION*, Jan. 2005, pp. 388–393.

[21] A. Bevilacqua, L. D. Stefano, and A. Lanza, "Coarse-to-fine strategy for robust and efficient change detectors," in *Proc. IEEE Conf. Adv. Video Signal Based Surveill.*, Sep. 2005, pp. 87–92.

[22] D. Rodriguez-Fernandez, D. L. Vilarino, and X. M. Pardo, "CNN implementation of a moving object segmentation approach for real-time video surveillance," in *Proc. Int. Workshop Cellular Neural Netw. Appl.*, Jul. 2008, pp. 129–134.

[23] Á. Bayona, J. C. SanMiguel, and J. M. Martínez, "Stationary foreground detection using background subtraction and temporal difference in video surveillance," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 4657–4660.

[24] J. Kim, B. Kang, H. Wang, and D. Kim, "Abnormal object detection using feedforward model and sequential filters," in *Proc. IEEE Int. Conf. Adv. Video Signal-Based Surveill.*, Sep. 2012, pp. 70–75.

[25] R. Mathew, Z. Yu, and J. Zhang, "Detecting new stable objects in surveillance video," in *Proc. IEEE Workshop Multimedia Signal Process.*, Oct./Nov. 2005, pp. 1–4.

[26] L. Maddalena and A. Petrosino, "Stopped object detection by learning foreground model in videos," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 5, pp. 723–735, May 2013.

[27] R. Wang, F. Bunyak, G. Seetharaman, and K. Palaniappan, "Static and moving object detection using flux tensor with split Gaussian models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2014, pp. 420–424.

[28] S. Guler and M. K. Farrow, "Abandoned object detection in crowded places," in *Proc. PETS*, 2006, pp. 18–23.

[29] K. Ingersoll, P. C. Niedfeldt, and R. W. Beard, "Multiple target tracking and stationary object detection in video with recursive-RANSAC and tracker-sensor feedback," in *Proc. Int. Conf. Unmanned Aircraft Syst.*, Jun. 2015, pp. 1320–1329.

[30] A. Bevilacqua and S. Vaccari, "Real time detection of stopped vehicles in traffic scenes," in *Proc. IEEE Conf. Adv. Video Signal Based Surveill.*, Sep. 2007, pp. 266–270.

[31] K. C. Smith, P. Quelhas, and D. Gatica-Perez, "Detecting abandoned luggage items in a public space," IDIAP, Idiap Res. Inst. Martigny, Switzerland, Tech. Rep.IDIAP-RR 06-39, 2006.

[32] S. Denman, V. Chandran, and S. Sridharan, "Abandoned object detection using multi-layer motion detection," in *Proc. Int. Conf. Signal Process. Commun. Syst.*, 2007, pp. 439–448.

[33] J.-Y. Chang, H.-H. Liao, and L.-G. Chen, "Localized detection of abandoned luggage," *EURASIP J. Adv. Signal Process.*, no. 1, Jun. 2010, Art. no. 675784.

[34] H.-H. Liao, J.-Y. Chang, and L.-G. Chen, "A localized approach to abandoned luggage detection with foreground-mask sampling," in *Proc. IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Sep. 2008, pp. 132–139.

[35] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Jun. 1999, p. 252.

[36] Y. L. Tian, M. Lu, and A. Hampapur, "Robust and efficient foreground analysis for real-time video surveillance," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1. Jun. 2005, pp. 1182–1187.

[37] J. L. Raheja, C. Malireddy, A. Singh, and L. Solanki, "Detection of abandoned objects in real time," in *Proc. Int. Conf. Electron. Comput. Technol.*, vol. 2. Apr. 2011, pp. 199–203.

[38] Y. Tian, R. S. Feris, H. Liu, A. Hampapur, and M.-T. Sun, "Robust detection of abandoned and removed objects in complex surveillance videos," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 41, no. 5, pp. 565–576, Sep. 2011.

[39] A. Lopez-Mendez, F. Monay, and J.-M. Odobez, "Exploiting scene cues for dropped object detection," in *Proc. Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, 2014, pp. 14–21.

[40] F. Porikli, "Detection of temporarily static regions by processing video at different frame rates," in *Proc. IEEE Conf. Adv. Video Signal Based Surveill.*, Sep. 2007, pp. 236–241.

[41] X. Li, C. Zhang, and D. Zhang, "Abandoned objects detection using double illumination invariant foreground masks," in *Proc. ICPR*, Aug. 2010, pp. 436–439.

[42] W. Wang and Z. Liu, "A new approach for real-time detection of abandoned and stolen objects," in *Proc. Int. Conf. Elect. Control Eng.*, Jun. 2010, pp. 128–131.

[43] U. A. Joglekar, S. B. Awari, S. B. Deshmukh, D. M. Kadam, and R. B. Awari, "An abandoned object detection system using background segmentation," *Int. J. Eng. Res. Technol.*, vol. 3, no. 1, pp. 1–4, Jan. 2014.

[44] L. Xiya, W. Jingling, and Z. Qin, "An abandoned object detection system based on dual background and motion analysis," in *Proc. Int. Conf. Comput. Sci. Service Syst.*, Aug. 2012, pp. 2293–2296.

[45] T. T. Zin, P. Tin, T. Toriu, and H. Hama, "A probability-based model for detecting abandoned objects in video surveillance systems," in *Proc. World Congr. Eng.*, vol. 2. 2012, pp. 1246–1251.

[46] A. Filonenko, Wahyono, and K.-H. Jo, "Detecting abandoned objects in crowded scenes of surveillance videos using adaptive dual background model," in *Proc. Int. Conf. Human Syst. Interactions*, Jun. 2015, pp. 224–227.

[47] Q. Li, Y. Mao, Z. Wang, and W. Xiang, "Robust real-time detection of abandoned and removed objects," in *Proc. Int. Conf. Image Graph.*, Sep. 2009, pp. 156–161.

[48] A. Singh, S. Sawan, M. Hanmandlu, V. K. Madasu, and B. C. Lovell, "An abandoned object detection system based on dual background segmentation," in *Proc. IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Sep. 2009, pp. 352–357.

[49] S. Cheng, X. Luo, and S. M. Bhandarkar, "A multiscale parametric background model for stationary foreground object detection," in *Proc. IEEE Workshop Motion Video Comput.*, Feb. 2007, p. 18.

[50] R. H. Evangelio and T. Sikora, "Static object detection based on a dual background model and a finite-state machine," *EURASIP J. Image Video Process.*, vol. 2011, no. 1, 2010. Art. no. 858502.
[51] S. Kwak, G. Bae, and H. Byun, "Abandoned luggage detection using a finite state automaton in surveillance video," *Opt. Eng.*, vol. 49, no. 2, p. 027007, Mar. 2010.
[52] Q. Fan and S. Pankanti, "Modeling of temporarily static objects for robust abandoned object detection in urban surveillance," in *Proc. IEEE Int. Conf. Adv. Video Signal-Based Surveill.*, Aug./Sep. 2011, pp. 36–41.
[53] Q. Fan and S. Pankanti, "Robust foreground and abandonment analysis for large-scale abandoned object detection in complex surveillance videos," in *Proc. IEEE Int. Conf. Adv. Video Signal-Based Surveill.*, Sep. 2012, pp. 58–63.
[54] A. Collazos, D. Fernández-López, A. S. Montemayor, J. J. Pantrigo, and M. L. Delgado, "Abandoned object detection on controlled scenes using Kinect," in *Natural and Artificial Computation in Engineering and Medical Applications*. Springer, 2013, pp. 169–178.
[55] C. Cuevas and N. Garcia, "Efficient moving object detection for lightweight applications on smart cameras," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 1, pp. 1–14, Jan. 2013.
[56] A. Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis, "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance," *Proc. IEEE*, vol. 90, no. 7, pp. 1151–1163, Jul. 2002.
[57] C. Cuevas, D. Berjón, F. Morán, and N. García, "Moving object detection for real-time augmented reality applications in a GPGPU," *IEEE Trans. Consum. Electron.*, vol. 58, no. 1, pp. 117–125, Feb. 2012.
[58] N. Martel-Brisson and A. Zaccarin, "Unsupervised approach for building non-parametric background and foreground models of scenes with significant foreground activity," in *Proc. ACM Workshop Vis. Netw. Behavior Anal.*, 2008, pp. 93–100.
[59] T. L. Booth, *Sequential Machines and Automata Theory*, vol. 3. New York, NY, USA: Wiley, 1967.
[60] D. Thirde, L. Li, and F. Ferryman, "Overview of the pets2006 challenge," in *Proc. IEEE Int. Workshop Perform. Eval. Tracking Surveill.*, Jun. 2006, pp. 47–50.
[61] i-LIDS Team, "Imagery library for intelligent detection systems (i-LIDS); a standard for testing video based detection systems," in *Proc. IEEE Carnahan Conf. Secur. Technol.*, Oct. 2006, pp. 75–80.
[62] C. Cuevas, E. M. Yáñez, and N. García, "Labeled dataset for integral evaluation of moving object detection algorithms: LASIESTA," *Comput. Vis. Image Understand.*, vol. 152, pp. 103–117, Nov. 2016.
[63] N. Goyette, P.-M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, "Changedetection.net: A new change detection benchmark dataset," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2012, pp. 1–8.
[64] C. Cuevas and N. García, "Improved background modeling for real-time spatio-temporal non-parametric moving object detection strategies," *Image Vis. Comput.*, vol. 31, no. 9, pp. 616–630, Sep. 2013.
[65] S. Boragno, B. Boghossian, J. Black, D. Makris, and S. Velastin, "A DSP-based system for the detection of vehicles parked in prohibited areas," in *Proc. IEEE Conf. Adv. Video Signal Based Surveill.*, Sep. 2007, pp. 260–265.
[66] S. Guler, J. A. Silverstein, and I. H. Pushee, "Stationary objects in multiple object tracking," in *Proc. IEEE Conf. Adv. Video Signal Based Surveill.*, Sep. 2007, pp. 248–253.
[67] P. L. Venetianer, Z. Zhang, W. Yin, and A. J. Lipton, "Stationary target detection using the objectvideo surveillance system," in *Proc. IEEE Conf. Adv. Video Signal Based Surveill.*, Sep. 2007, pp. 242–247.
[68] F. Porikli, Y. Ivanov, and T. Haga, "Robust abandoned object detection using dual foregrounds," *EURASIP J. Adv. Signal Process.*, Jan. 2008, Art. no. 197875.
[69] J. Pan, Q. Fan, and S. Pankanti, "Robust abandoned object detection using region-level analysis," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 3597–3600.

Carlos Cuevas received the Ingeniero de Telecomunicación degree (integrated B.Sc. and M.Sc. accredited by ABET) and the Doctor Ingeniero de Telecomunicación degree (Ph.D. in Communications) in 2011 (Doctoral Graduation Award) from the Universidad Politécnica de Madrid (UPM), Madrid, Spain, in 2006 and 2011, respectively. Since 2006, he has been a member of the Grupo de Tratamiento de Imágenes (Image Processing Group) of the UPM. Since 2013, he has been a member of the UPM's faculty. His research interests include signal and image processing, computer vision, pattern recognition, video coding, and automatic target recognition.

Raquel Martínez received the Ingeniero de Telecomunicación degree (integrated B.Sc. and M.Sc. accredited by ABET) from the Universidad Politécnica de Madrid, Madrid, Spain, in 2015. She has been collaborating with the Grupo de Tratamiento de Imágenes (Image Processing Group) in the UPM since 2014. Her research interests include signal and image processing and computer vision.

Daniel Berjón received the Ingeniero de Telecomunicación degree (integrated B.Sc. and M.Sc. accredited by ABET) and the Doctor Ingeniero de Telecomunicación degree (Ph.D. in Communications) from the Universidad Politécnica de Madrid (UPM), Spain, in 2006 and 2011, respectively. Since 2008, he has been a member of Grupo de Tratamiento de Imágenes (Image Processing Group). His research interests include image processing, parallel processing, computer graphics, and real-time systems.

Narciso García received the Ingeniero de Telecomunicación degree (five years engineering program) (Spanish National Graduation Award) and the Doctor Ingeniero de Telecomunicación degree (PhD in Communications) (Doctoral Graduation Award) from the Universidad Politécnica de Madrid (UPM), Madrid, Spain, in 1976 and 1983, respectively. Since 1977, he has been a member of the faculty of the UPM, where he is currently a Professor of Signal Theory and Communications. He leads the Grupo de Tratamiento de Imágenes (Image Processing Group), UPM. He has been actively involved in Spanish and European research projects, also serving as an Evaluator, a Reviewer, an Auditor, and an Observer of several research and development programs of the European Union. He was a Co-Writer of the EBU proposal, base of the ITU standard for digital transmission of TV at 34-45 Mb/s (ITU-T J.81). He was an Area Coordinator of the Spanish Evaluation Agency from 1990 to 1992 and he was the General Coordinator of the Spanish Commission for the Evaluation of the Research Activity (CNEAI) from 2011 to 2014. His current research interests include digital image and video compression and computer vision.