

Article

Background Subtraction Approach based on Independent Component Analysis

Hugo Jiménez-Hernández

Centro de Investigación en Ciencia Aplicada y Tecnología Aplicada Cerro Blanco No. 141. Col. Colinas del Cimataro, Santiago de Querétaro, Querétaro, Mexico

* Author to whom correspondence should be addressed; E-Mail: hugojh@gmail.com; Tel.: +52-1-442-229-0804; +52-1-442-211-9800 Ext. 1330; Fax: +52-1-442-211-9839.

Received: 26 April 2010; in revised form: 16 May 2010 / Accepted: 28 May 2010 /

Published: 18 June 2010

Abstract: In this work, a new approach to background subtraction based on independent component analysis is presented. This approach assumes that background and foreground information are mixed in a given sequence of images. Then, foreground and background components are identified, if their probability density functions are separable from a mixed space. Afterwards, the components estimation process consists in calculating an unmixed matrix. The estimation of an unmixed matrix is based on a fast ICA algorithm, which is estimated as a Newton-Raphson maximization approach. Next, the motion components are represented by the mid-significant eigenvalues from the unmixed matrix. Finally, the results show the approach capabilities to detect efficiently motion in outdoors and indoors scenarios. The results show that the approach is robust to luminance conditions changes at scene.

Keywords: background subtraction; independent component analysis; motion detection

1. Introduction

One of fundamental steps, in several computer vision systems, is the motion detection process. The moving objects represent the main features used to analyze the motion dynamics at scene. The background and foreground commonly represent fixed areas and moving areas respectively. The process of identifying foreground and background objects is a tough task. The background subtraction approach

consists on label moving objects and fixed regions. However, there are several factors, like luminance, reflections, shadows, or even camera shaking that make this process difficult [1, 2].

The foreground labeling process can be considered as a general classification problem, where using a model M , we try to estimate a set of parameters $P = \{p_1, p_2, \dots\}$, which correctly label/classify background and foreground objects. The parameter estimation uses previous knowledge about scenario and object properties. The capabilities of classification depend on both the raw data and the model M . It is usually pretended to build a single binary classifier [2–4] (if it only identifies fixed and moving zones) or multiple classifiers [1, 5] (if it wishes to model several types of moving objects and fixed layers). The success of motion detection for a particular model depends on the scene constraints, the data dynamics on both temporal and spatial conditions, and the separability of data under the current model M . In the literature, one of the most accepted approaches is the work of Stauffer and Grimson [3]. They proposed an approach based on the assumption that the foreground and the background can be modeled as a mixture of Gaussians (MOG), where the Gaussians with low probability represent moving objects. This approach is computationally efficient; moreover, the convergence velocity and the spatial object consistency are not considered. Furthermore, in their work, Toyama *et al.* [1] proposed a background approach capable of discarding periodic movements in background as the waves of the sea or the up and down of electric stairs. Moreover, this approach is limited to scenarios with the fixed luminance conditions. Next, Elgammal *et al.* [6] proposed an approach based on a pixel probability density function (pdf) approximation using last immediate frames, which are smoothed with a Gaussian kernel. Also, Horprasert *et al.* [2] proposed a method based on object segmentation that supports shadows and changing luminance conditions, against an increase of the computational complexity. The same form, Oliver *et al.* [7] proposed a background approach based on a temporal correlation of pixel values. Next, the Principal Component Analysis (PCA) [8] is applied to eliminate those components, which do not provide information to the model. This approach was improved to support small luminance changes by Rymel *et al.* [9]. More recently, Han *et al.* [10] presented an approach to estimate a set of models of pixel intensities based on a combination of mean-shift pdf's estimators and a propagated Gaussian kernel. Following the trend, Tang and Miao [11] presented an extension to the MOG approach, which supports shadows. Other relevant work is proposed by Zhou *et al.* [12], where they show a different background approach, based on the analysis of image texture using Gabor filters. Finally, at a recent time, in their work [13], Du-Ming and Shiah-Chin proposed a way to detect motion based on Independent Component Analysis (ICA), which is used to estimate the background information. This approach is based on the Particle Swarm Optimization (PSO) algorithm to search the best unmixing matrix. The background process consists on estimate the unmixing matrix using a background estimation and the current image. Afterwards, the approach, for consecutive frames, separates the background and the foreground objects as a single threshold classification task, using the same estimated unmixed matrix. Moreover, this approach is limited to scenarios where the background dependences between foreground and background never change. As a consequence, in outdoor scenarios this approach is not suitable.

This work presents a novel background subtraction approach based on Independent Component Analysis [14, 15]. This approach exploits the property of separability of the pdf in several components; in which one of them represents the background, and the rest of the components represent foreground

areas and noise effects. This approach considers the task as a cocktail problem, where background and foreground information are mixed spatially and temporally. The process consists on identifying the pdf of the background from the rest, separating the different components as a linear independent component. The separation process is performed with an unmixing matrix. The parameters of the unmixing matrix are estimated as a maximization of non-Gaussian problem, which is solved via Newton-Raphson[16]. Background and foreground zones are detected by extracting the most and middle significant components from the unmixed matrix. To compensate the dynamic changes of the scenario, the approach is continuously estimating the unmixed matrix and the estimation of background information. As a consequence, this approach is adaptable to different luminance conditions. The paper is organized as follows. In §2, a background model based on ICA is presented. The parameter estimation for the unmixing matrix is discussed in §3. Next, the motion detection from the unmixed matrix is presented in §4. Afterwards, in §5 the approach is tested, comparing foreground detection with the MOG approach [3]. Both approaches are tested in several image sequences took from a PETS database [17, 18] and outdoors/indoors scenarios (gardens and vehicular intersections). Using the PETS database we can develop a background bench scheme to quantify the accuracy of tested approaches. Finally, the conclusion is shown.

2. Background Model

Given a set of consecutive images $\Upsilon = \{I_1, I_2, \dots, I_n\}$, the information of moving objects and fixed objects are mixed. Each particular image pixel position $\mathbf{x} = [x_1, x_2]$ is indexed as $I_i(\mathbf{x})$ for $k \times l$ image dimensions. To start with, an $E(\Upsilon)$ operator is defined. This operator mixes the images in Υ and estimates the dominant color pixel value for each position \mathbf{x} . Table 1 shows some common operators used for this purpose, even though, there are many others operators; the complexity and computational resources could increase. The estimation of the color of the pixel depends on the temporal window size used. For reducing the computational complexity, the estimation is approximated with recursive algorithms based on time differences, *i.e.*, Kalman recursive filters [19], or parameter estimation using Expectation Maximization algorithm [20].

Table 1. List of some common function estimator $E(\Upsilon)$ for pixel values.

No.	Estimator
1	$E(\Upsilon) = \tilde{I}(\mathbf{x})$ where each $\tilde{I}(\mathbf{x}) = \frac{1}{\ \Upsilon\ } \sum_{i \in \Upsilon} I_i(\mathbf{x})$
2	$E(\Upsilon) = \tilde{I}(\mathbf{x})$ where each $\tilde{I}(\mathbf{x}) = \text{median}\{I_{i \in \Upsilon}(\mathbf{x})\}$
3	$E(\Upsilon) = \mu_{\mathbf{x}}$ where each $\tilde{I}(\mathbf{x}) \sim G(\mu_{\mathbf{x}}, \sigma_{\mathbf{x}})$

This work assumes that motion objects and background areas can be represented as independent image components U_1, U_2, \dots . The most significant component represents the background image, and the rest of the components represent moving objects, and the less significant may represent noise motion. But, these images are unfortunately, unknown. Using $E(\Upsilon)$ and a set of images Υ , the estimation of the matrix W is performed. This matrix unmixes into a set of images U_1, U_2, \dots , which represent the independent components of fixed and moving objects. Namely, for a particular position \mathbf{x} we have,

$$\Phi(\mathbf{x}) = W\Omega(\mathbf{x}) \quad (1)$$

where the W matrix separates the mixed color components in $\Omega(\mathbf{x}) = [I_1(\mathbf{x}), I_2(\mathbf{x}), \dots, I_n(\mathbf{x}), E(\Upsilon(\mathbf{x}))]$; $\Phi(\mathbf{x})$ represents the unmixed independent components $\Phi(\mathbf{x}) = [U_1(\mathbf{x}), U_2(\mathbf{x}), \dots, U_n(\mathbf{x}), U_{n+1}(\mathbf{x})]$. The estimation of W needs to express each image in Υ as a vector. Then, this transformation is performed with $v(I_i)$ transformation. This transformation catches spatial information of texture and luminance conditions, expressing any image I_i as a vector form \mathbf{I}_i . The transformation $v(I_i)$ is useful if we need to reduce the computer complexity when images are too big, or when there are some restrictions of zones scene. Additionally, the transformation $v(I_i)$ implies that the inverse exists $v^{-1}(\mathbf{I}_i)$, that is, the images encoded as vectors \mathbf{I}_i can be mapped into original images again, preserving the same structure than the original images encoded I_i . For the rest of the document, the bold letter images refer to the vector version resulted from applying v transformation. Next, the parameter estimation of the matrix W is needed for detecting moving objects from an image sequence. Both processes are discussed in the following sections.

3. Estimation of Unmixing Matrix

The matrix W is estimated assuming that the pdf of each $\mathbf{U}_i(\mathbf{x})$ are both separable and independent from the mixed matrix $\Omega = [\mathbf{I}_1, \dots, \mathbf{I}_n, \mathbf{I}_{n+1}]$ for all $\mathbf{I}_i \in \Upsilon$ for $n = 1, \dots, n$ and $\mathbf{I}_{n+1} = E(\Upsilon)$, where $E(\Upsilon)$ represents the image estimator and Υ raw data, *i.e.*, the joint probability of all pdfs are factored as follows

$$p(\mathbf{U}_1, \dots, \mathbf{U}_{n+1}) = p(\mathbf{U}_1)p(\mathbf{U}_2) \dots p(\mathbf{U}_{n+1}) \quad (2)$$

where $p(\mathbf{U}_i)$ represents the pdf of unmixed images expressed as vector and the independent components are represented in matrix form as $\Phi = [\mathbf{U}_1, \dots, \mathbf{U}_n, \mathbf{U}_{n+1}]$. The parameter estimation is performed by identifying data directions that decrease the Gaussianity of the mixed distributions. In this sense, we can assume that the pdf of the moving objects \mathbf{u}_i has a non-Gaussian distribution. The background distribution behaves mainly as a Gaussian, and it does not affect the component separation process, whenever the other components become non-Gaussian. This assumption establishes the minimum criterion to separate the independent components from the mixed data. In case that all pdfs have a Gaussians distribution, the mixed distribution is symmetrical and is not possible to separate it, *i.e.*, if objects with motion have the same pdf as background, they could have not been identified.

3.1. Image Preprocessing

In a preprocessing step, all mixed images \mathbf{I}_i are centered subtracting the mean value $m(\mathbf{I}_i)$ so as to make a zero-mean variable. After, each \mathbf{I}_i is uncorrelated with a linear transformation for which its components are equal to one, as follows,

$$\tilde{\mathbf{I}}_i = \mathbf{S}\Sigma^{-1/2}\mathbf{S}^T\mathbf{I}_i \quad (3)$$

where $\tilde{\mathbf{I}}_i$ denotes uncorrelated version of \mathbf{I}_i ; Σ and \mathbf{S} result from factorizing covariance matrix $E\{\mathbf{I}_i\mathbf{I}_i^T\} = \mathbf{S}\Sigma\mathbf{S}^T$. Finally, $\Sigma^{-1/2}$ is computed by a simple component-wised operation.

3.2. Measure of Gaussianity

Then, since data are uncorrelated, the weight parameters of $W = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n]^T$ are estimated for each row \mathbf{w}_i such that the projection $\mathbf{w}_i^T \tilde{\mathbf{I}}$ maximizes the non-Gaussianity. In this sense, the non-Gaussianity degree is measured with an approximation of neg-entropy [14]. The advantages to use neg-entropy instead of kurtosis, for instance, are that is well justified in statistical theory and in some sense, neg-entropy is the optimal estimator of non-Gaussianity as far as statistical properties are concerned. Therefore, a simpler approximation of neg-entropy could be estimated as follows,

$$J(y) \propto [E\{G(y)\} - E\{G(v)\}]^2 \quad (4)$$

where G is commonly a non-quadratic function. The election of any approximation of G depends of the scene conditions and the behavior of raw data; however, there are some common approximations of neg-entropy that are shown in Table 2.

Table 2. Common estimator functions G for a neg-entropy approximation.

No.	Function
1	$G(u) = \frac{1}{a_1} \log \cosh a_1 u$ for $0 \leq a_1 \leq 1$
2	$G(u) = -e^{-\frac{u^2}{2}}$
3	$G(u) = u^3$

3.3. Parameter Estimation

At this point, the task to estimate the W weight values, is considered as an optimization problem. In this sense, using an approach based on Newton-Raphson like [16], an optimization procedure based on fast ICA [14] is performed to estimate each of the \mathbf{w}_i weight vector parameters in W as follows:

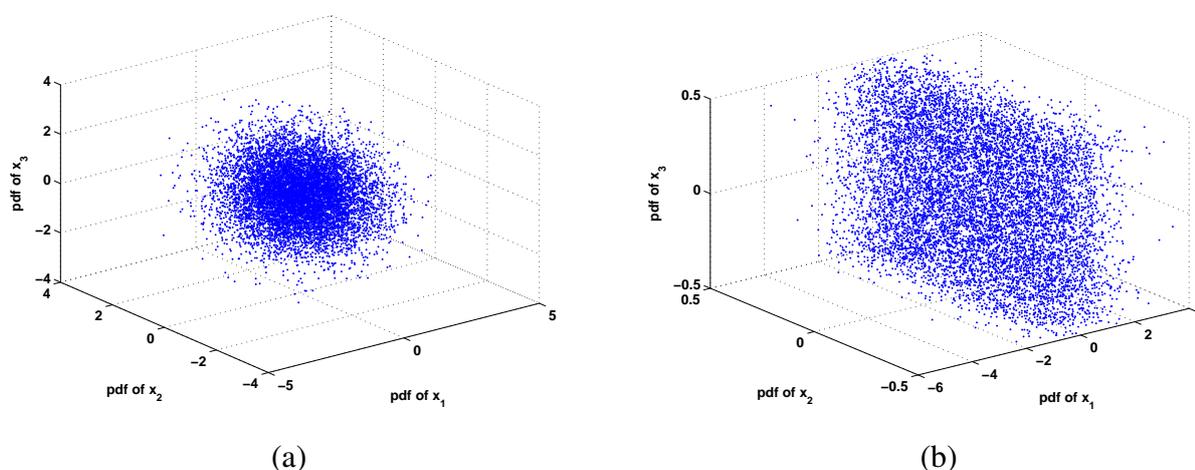
1. For each \mathbf{w}_i in W
 - (a) Choose initial (*i.e.*, random) values weight for \mathbf{w}_i .
 - (b) Let $\mathbf{w}_i^+ = E\{\tilde{\mathbf{I}}_i g(\mathbf{w}_i^T \tilde{\mathbf{I}}_i)\} - E\{g'(\mathbf{w}_i^T \tilde{\mathbf{I}}_i)\} \mathbf{w}_i$.
 - (c) Let $\mathbf{w}_i^+ = \frac{\mathbf{w}_i^+}{\|\tilde{\mathbf{I}}_i^+\|}$.
 - (d) If convergence is not achieved, go back to 1 (a).
2. Continue with the next \mathbf{w}_i

Each internal iteration only estimates one \mathbf{w}_i vector. Then, the algorithm is repeated $n + 1$ times, once by each image analyzed in Ω . In [21] is shown a parallel approach to this algorithm, with the advantage of time reduction, against of a considerable reduction in the precision level [22].

3.4. The Background Gaussianity and Independent Component Separability

The matrix W is a linear transformation that separates each non-Gaussian pdf contained in an $I_i(\mathbf{x})$ image. But, Stauffer and Grimson [3] have proved experimentally that the background can be modeled as a Gaussian distribution. This implies that there is always one component that is Gaussian. The symmetrical morphology of a Gaussian distribution may affect the optimization steps 1(b) and 1(c), which assume non-Gaussian distribution over each component. In the practice, this constraint can be relaxed, considering the effects that this implies, *i.e.*, if we had all the components distributed as a Gaussian, these could be represented as a hipper ball; which we have not gotten enough geometrical information to separate each independent component (because they are totally symmetrical and they do not have any maximum). But when one or some components are Gaussians, the non-Gaussian components add vertexes (*i.e.*, maximums), which are used to estimate each independent component. Figure 1 illustrates the space conformed by three different pdfs; whenever they follow a Gaussian distribution (Figure 1a) there is not enough information for unmixing, instead, when only one has a Gaussian distribution, the others provide geometrical information for unmixing the space. As it is appreciated, the background Gaussianity does not affect the unmixed process at the parameter estimation process, whenever the pdfs of moving objects have non-Gaussian distribution. This situation may be appreciated as a limitation, but in practice, the majority of moving objects, in short time stamps, does not follow a Gaussian distribution, thus being the only background component considered having Gaussian distribution.

Figure 1. Data distribution for 3 variables (a) when data have a Gaussian Distribution there is not enough information to estimate each independent component (the mixed distribution is totally symmetrical); (b) when data have mixture of non-Gaussian or some of the independent components are Gaussian, it is possible to identify other non-Gaussian components (there are some vortex that define maximums).



4. Motion Detection

The data in Equations 1 and 2 are uncorrelated and normalized, being impossible to rank or define an order over all the estimated components. However, the values contained at W provide information about the amount of information over the estimated components. The matrix W represents the linear transformation that separates the mixed images as independent components, the inverse of the transformation W^{-1} mixes up again the sources $U(\mathbf{x})$. The linearity of each component expressed in W is used to define the importance of each component and, as a consequence, the detection of background, foreground and noise components. The difference between using a PCA [8] approach instead of an ICA, consists in the extra constraint added to the component estimation; *i.e.*, the components must be orthogonal and uncorrelated at same time. Next, from analysis of the singular values of W^{-1} , the most significant singular value component represents the background data, under the assumption that fixed areas are proportionally greater than moving areas in scene. The next most significant singular values (except for the first one) correspond to moving components, and the last components, represent noise motion.

Then, to separate the first (background component) and last components (noise components) on W^{-1} matrix, the singular value decomposition is applied erasing the eigenvalues that correspond to the most significant component and the less significant components. Rebuilding the matrix W^{-1} without the first one and the last eigenvalues, the components Ω^* are estimated as follows

$$\Omega_i^* = W^{*-1}\Phi = (S\Sigma^*D^T)\Phi \quad (5)$$

where $W^{-1} = S\Sigma D^T$; $\Sigma^* = \Sigma$ without the most significant and the less significant components; *i.e.*, $\Sigma_{11}^* = 0$ and $\Sigma_{ii} = 0$ for $\{i, i+1, \dots, n\}$, which preserves the same information except for the background and noise information; and $W^{*-1} = S\Sigma^*D^T$. The data, in the reconstructed image Ω^* , contain only data corresponding to the moving objects of each independent component U_i contained in Φ . Next, it is needed to binarize each one of $I_i^* = v^{-1}(\mathbf{I}_i^*)$ that conforms $\Omega^* = [\mathbf{I}_1^{*T}, \mathbf{I}_2^{*T}, \dots, \mathbf{I}_n^{*T}, \mathbf{I}_{n+1}^{*T}]^T$ for finding out moving objects.

Each particular image I_i^* does not contain background data and noise data; *i.e.*, the effect of putting $\Sigma_{11} = 0$ and some $\Sigma_{ii} = 0$ consists in that the information over each component has been discarded, resulting in spurious data around zero value. In fact, the background corresponds to the majority of the area, and it is represented by the pdf global maximum, and the moving objects by the rest of the pdfs local maximum. The values, that belong to the biggest maximum, are labeled with zero value (which is located around zero value), and rest to one. The resulting binary map represents moving objects as follows

$$B(\mathbf{x}) = \begin{cases} 0 & \text{if } I_i^*(\mathbf{x}) \in \text{Background} \\ 1 & \text{other case} \end{cases} \quad (6)$$

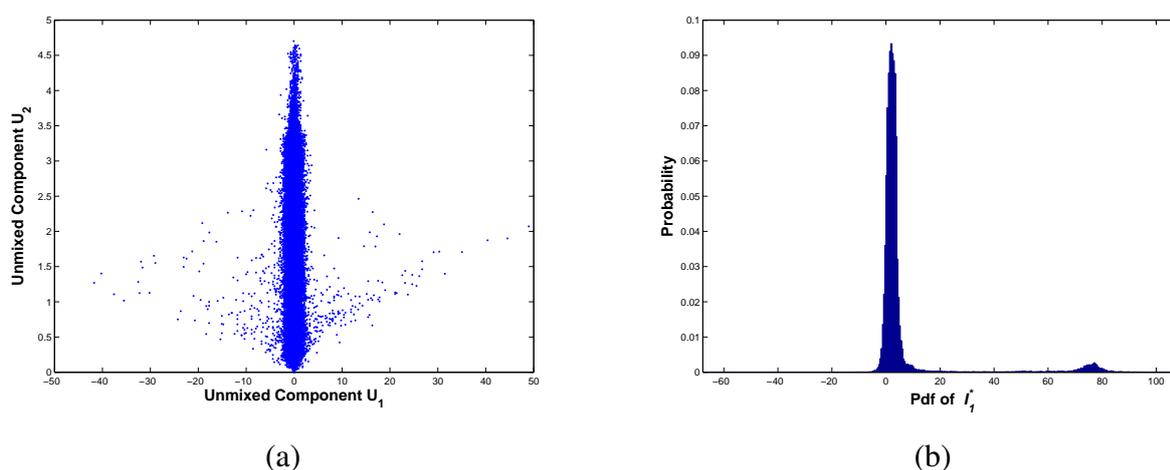
where the labeling process is achieved via a k-means's variation [23], which group all the pixels in two categories; background and foreground.

As an example, using the estimated image $E(\Upsilon)$ and the current image I_t , Figure 2a shows the distribution of independent unmixed components U_1 and U_2 , after estimating the unmix matrix W . The main component represents the background information (vertical component), and lower components

represent the moving objects information (horizontal component). Then, the reconstructed I_t^* is estimated from the mixing matrix W^{*-1} . The pdf of I_t^* is presented in Figure 2b. The data that surround the zero value correspond to the background; the rest corresponds to the moving components. Finally, the classification task consists on grouping all the elements around zero labeling them as background, and the rest as moving objects.

The motion detection process is invariant to the global luminance variations and diffuse lights. The global luminance variations are tolerated, in sense, they represent the displacement of the pdf along its range. As a consequence, the uncorrelated process always centers and normalizes raw data, being not affected thus by the displacements on range. The diffuse lights affect locally the raw data, but do not affect considerably the pdf distribution of the pixel. Then, this does not have a significant effect in the independent component separation process. The luminance invariants make feasible that this approach supports blur shadows, whenever the pdf morphology is not affected (experimental evidence is presented in §5). Moreover, motion detection may be degenerated when the parameter estimation does not converge, avoiding the correct detection. This happens when both foreground and background have Gaussian distribution, or foreground raw data has not enough evidence to, numerically, estimate its pdf.

Figure 2. Foreground detection from independent components using the estimation $E(\Upsilon)$ and current image I_t . In (a) the unmixed space as orthogonal components resulted from the pdfs of U_1 and U_2 ; the main component represents background information (vertical component) and the second component represents foreground objects; In (b) the pdf of estimated I_t^* after removing background data. The global maxima correspond to background data, and second maxima (right side) correspond to objects with motion.



As a complement, the component estimation of moving objects could be affected with noise. An additional step consists in applying a connectivity analysis. It can be performed with a morphological operator P applied to $B(\mathbf{x})$ [24], where an useful morphological filter is the opening by reconstruction, which removes low connected areas. The motion areas may be deformed, because, these operators assume that the structural element is not affected by the current image camera projection; *i.e.*, the background is fronto-parallel to the camera view. The opening by reconstruction filter removes elements low-connected than the structural element λ , resulting in a better connected map B . Then,

the morphological filtering process of the binary map is denoted as follows

$$\tilde{B} = \tilde{\gamma}(B) = \lim_{n \rightarrow \infty} \delta_B^n(\varepsilon_\mu(B)) \quad (7)$$

The filtered motion map \tilde{B} enhances the motion zones discarding additive noise that could affect the process.

5. Experiments and Results

In this section we present an experimental model to test our approach and some implementation issues. The experimental model matches the performance of motion detection between MOG [3] and our approach, against a ground truth based on PETS database [17, 18]. The implementation issues point out some useful remarks to reduce the complexity of motion detection.

5.1. Implementation Issues

To reduce the computational complexity of the algorithm, we should take the following modification to the approach. The $E(\Upsilon)$ operator should be based on a recursive efficient estimator to compute the expected value [19, 20]. But, we must consider that a simple model could affect the accuracy of the $E(\Upsilon)$ estimator. Additionally, the $v(I_i)$ function should sample each image, for avoiding the use of Gaussians filters, which in certain conditions are equivalent [25], but the sampling process reduces the computational resources needed. The number of independent components to be estimated in Equation(1) should be reduced to only the image estimator $E(\Upsilon)$ and the last image acquired. This makes that the unmixed matrix W becomes to a 2×2 dimension. Now, the second eigenvalue of Σ^* in Equation(5) represents the amount of information of moving objects and noise motion, being a necessary thresholding step over a second eigenvalue to detect object motion and noise motion.

5.2. Experimental Model

A quantitative process to measure the reliability of a background approach is not an easy task. In background model context, some works like [26] and [27] propose methods for quantifying the degree of success of different background approaches. Moreover, many times, background models are a dependable application, making these proposals not completely general. In this sense, we introduce a quantitative analysis based on the comparison of a well accepted background method as a reference and a set of sequences of images as motion ground truth. Both provided a relative point of comparison to evaluate the effectiveness and accuracy of motion detection in different scenarios.

A performance quantitative measure of our background model consists in comparing motion detection efficiency between our approach and a reference model. The MOG approach [3] is used as a reference model. This approach is a well accepted method and it generally offers good results. To measure the algorithm performance, a motion ground truth is introduced, which is conformed by several sequences of images taken from PETS database [17, 18]. PETS databases are used as a reference for evaluating surveillance algorithms. The sequences of images used are taken from 2001 and 2007 PETS databases, which include indoors/outdoors scenarios. The motion ground truth is performed by hand-selecting motion zones. Only the objects displaced are considered as motion zones, discarding reflections, noise

and shadows. We use, as ground truth, three different sequences (see Figure 3). Table 3 summarizes the principal information of sequences used as ground truth.

We introduce three different error measures to compare the two approaches against ground truth data. These errors are expressed in pixels. The first one quantifies the degree of modeling between the foreground estimated map and its ground truth map. This error is well known as BIAS error [28], which, in our context, consists in the arithmetical summation of a binary image difference between motion detection performed by the background approach and its ground truth. This measure provides information about the under/over modeled of motion scenario. Formally, this measure is defined as

$$\epsilon_b = \sum_{\mathbf{x} \in B} B(\mathbf{x}) - I_{gt}(\mathbf{x}) \quad (8)$$

where $B(\mathbf{x})$ is the foreground estimated map and $I_{gt}(\mathbf{x})$ is the ground truth map. The second measure, represents the number of pixels labeled as motion zones that correspond to free movement zones. It is commonly named as a false-positive error [29] and it is defined as the complement of a $I_{gt}(\mathbf{x})$ map multiplied point to point with a current motion map $B(\mathbf{x})$. The error measures near to zero, meaning that the background approach discards efficiently the free movement zones, avoiding thus, the addition of motion noise at $B(\mathbf{x})$ map. Formally, this error is denoted as follows,

$$\epsilon_{fp} = \sum_{\mathbf{x} \in B} \overline{B(\mathbf{x})} * I_{gt}(\mathbf{x}) \quad (9)$$

where $\overline{B(\mathbf{x})}$ is the complement of the map of motion detected $B(\mathbf{x})$ and the operator $*$ is the single multiplication. The third measure represents the number of pixels labeled as free motion zones that correspond to motion zones. This error is named as a false-negative error and represents the capabilities of our approach to model correctly motion zones. Error measures near to zero, mean that all objects with motion have been adequately detected. Consequently, error measures greater than zero, mean that the background approach does not detect efficiently the motion areas. This error measure is defined as follows

$$\epsilon_{fn} = \sum_{\mathbf{x} \in B} B(\mathbf{x}) * \overline{I_{gt}(\mathbf{x})} \quad (10)$$

where $\overline{I_{gt}(\mathbf{x})}$ is the complement of ground truth map and $*$ operator is single multiplication.

Next, the approach is tested with a set of sequences taken from different outdoor/indoor scenarios. These scenarios show different common situations in surveillance systems like luminance disturbance, shadows, reflection, *etc.*

Then, the work is compared with the PCA approach (see Appendix A) for segmenting the background from the foreground. Finally, it is shown the results for applying a morphological analysis to improve the quality of the motion map.

Table 3. List of sequences of images used as ground truth. These images present complex scenarios. The first two sequences represent scenarios with shadows caused by different light sources and both color and texture of skin in some frames, that are too similar to background scenario. The third one, represents a scenario with high noise degree, the reflections are caused by the windows and the moving objects are represented by small zones.

No.	Place	Source	Num. Img. ¹	bf Num. Train. ²
1	Train Station	PETS 2007	300	100
2	Train Station	PETS 2007	300	100
3	Outdoors Park	PETS 2001	500	100

¹ Total number of images used.

² Number of images used to train background model approach.

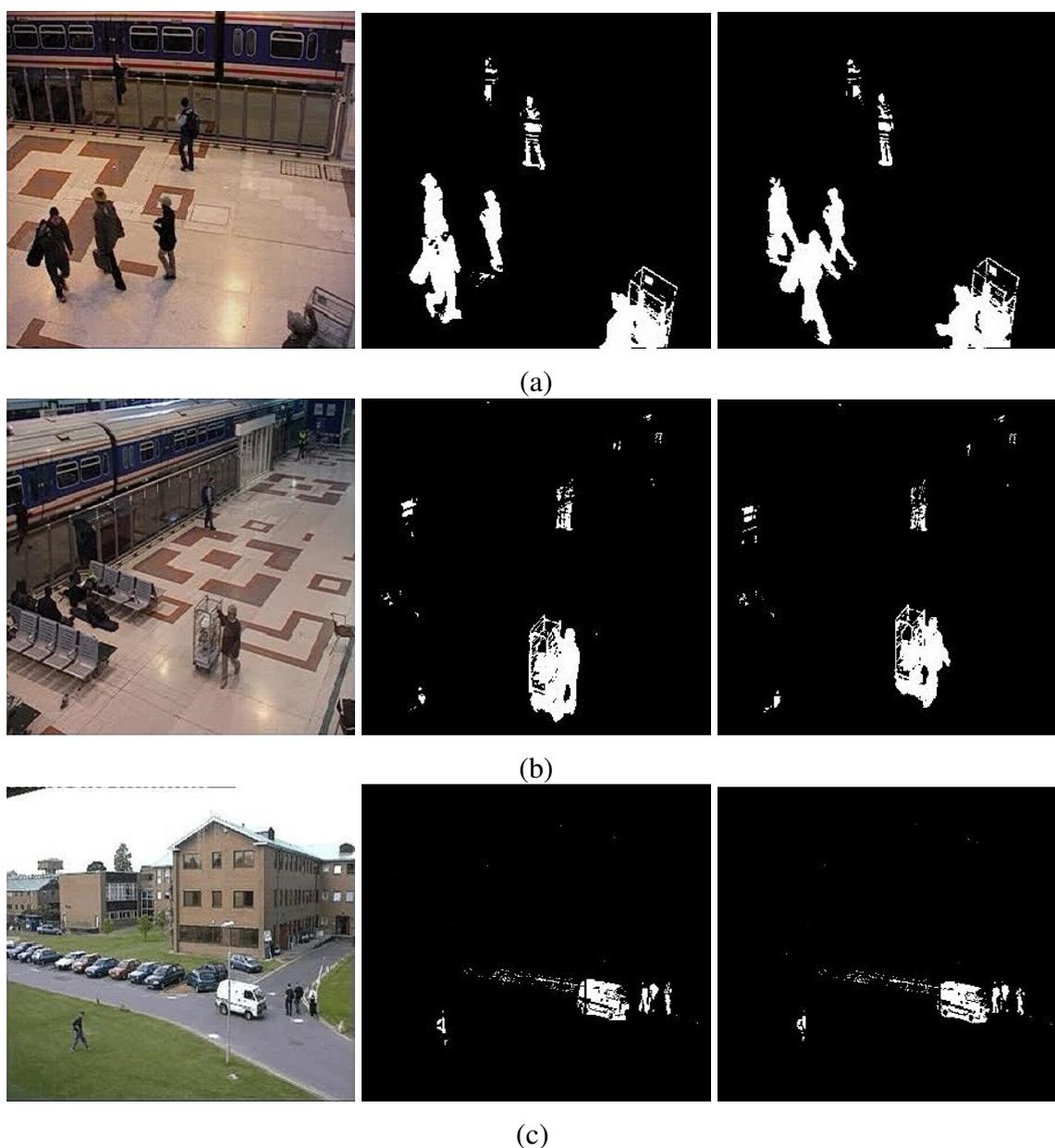
5.3. Results and Discussions

The proposal and MOG approach are tested as we described above. The proposal was implemented with only two components (the estimator of images $E(Y)$ and the last frame acquired), the function $v(I)$ is defined as the concatenation of odd columns conformed by odd pixel positions. The threshold to identify moving objects and noise is defined to 0.05. The operator $E(Y)$ is defined as an average operator to estimate background information with a window of 100 frames. The MOG approach was implemented using a $\rho = 0.005$, and 3σ as belonging criterion. The background initialization was performed with 100 frames. All tests are performed without applying any connectivity criterion.

The results of testing the proposal and the MOG approach are showed in Figures 4–6. Additionally, Figures 3 and 7 show some frame motion detected with both approaches. As it is appreciated, in Figure 4, the MOG approach is generally more sensitive to changing light conditions in the scene; on the other hand, the proposed approach becomes more stable in tested scenes. The bias error is greater with a MOG approach than the proposal, *i.e.*, the proposal detects better objects motion in the sequences of images.

The false-positive error quantifies when the approach adds noise from motion to the motion detection map. In this sense, we appreciate in Figure 5 that the MOG approach detects false motion zones. The false motion zones are difficult to deal with, and several times, it would be confusing with the motion objects, degrading the post analysis stages in vision systems. Mainly, the noise is caused by the sudden reflections and the shadows caused by the objects moving, like people walking down (Figure 7a), in MOG approach; but in the proposal, the error measure is small, behaving more efficiently than the MOG approach (Figure 3a). In outdoor scene, MOG error remains constant and usually greater than the proposed approach. Thus, the MOG approach produces binary maps of motion affected by noise, whereas the proposal produces binary maps of motion more clearly, without applying any connectivity analysis. The sudden noise peaks are caused particularly when objects are too small (a deep discussion is made in the following paragraphs).

Figure 3. Foreground detection from image sequences of the PETS database [17, 18] using the proposal. The sequences in (a) and (b) correspond to cameras monitoring a train stop; the sequence in (c) corresponds to a car parking. The foreground detection is performed without spatial filter. The level of resolution is higher and soft shadows are discarded automatically.



On the other hand, the false-negative error quantifies the accuracy to detect foreground areas. In this sense, we appreciate that the MOG approach usually detects correctly objects with motion. However, when foreground has a high similarity degree with background or it is constituted by big flat surfaces, the MOG approach only detects foreground contours. This fact causes a high false-negative error with the MOG approach (see Figure 6a–c). The proposal estimates and separates the mixed pdfs that conforms background and foreground zones. Consequently, it produces a better motion segmentation, and a small false-negative error. In Figure 6a both approaches behave too similar. The small variations are caused when skin zones belonging to people walking down are too similar with background. In Figure 6b, the camera perspective and the reflections difficult the process of foreground detections, causing that people

walking down are not detected. Finally, in outdoor scenarios, the small objects are detected efficiently with MOG; however, big objects with slow motion are broken (Figure 7c). But our approach segment with fewer noise degree, and the slow movement objects are better segmented (Figure 3c). Consequently, the error level is similar in both cases, except for the end of the sequence that corresponds to a vehicle with slow motion. Visually, the results can be appreciated in Figures 3 and 7.

Figure 4. Bias error resulted to compare ground truth against our approach and the MOG approach. In (a) it is shown the bias error of the first scenario that corresponds to a train station. In (b), the same train station is monitored from other perspective, where the reflection, the shadows, and the perspective make more complicated the motion detection. In (c) the error bias at an outdoors scenario; the noise in the sequence affects negatively the motion detection. The proposed approach generally offers a smaller bias error than the MOG approach.

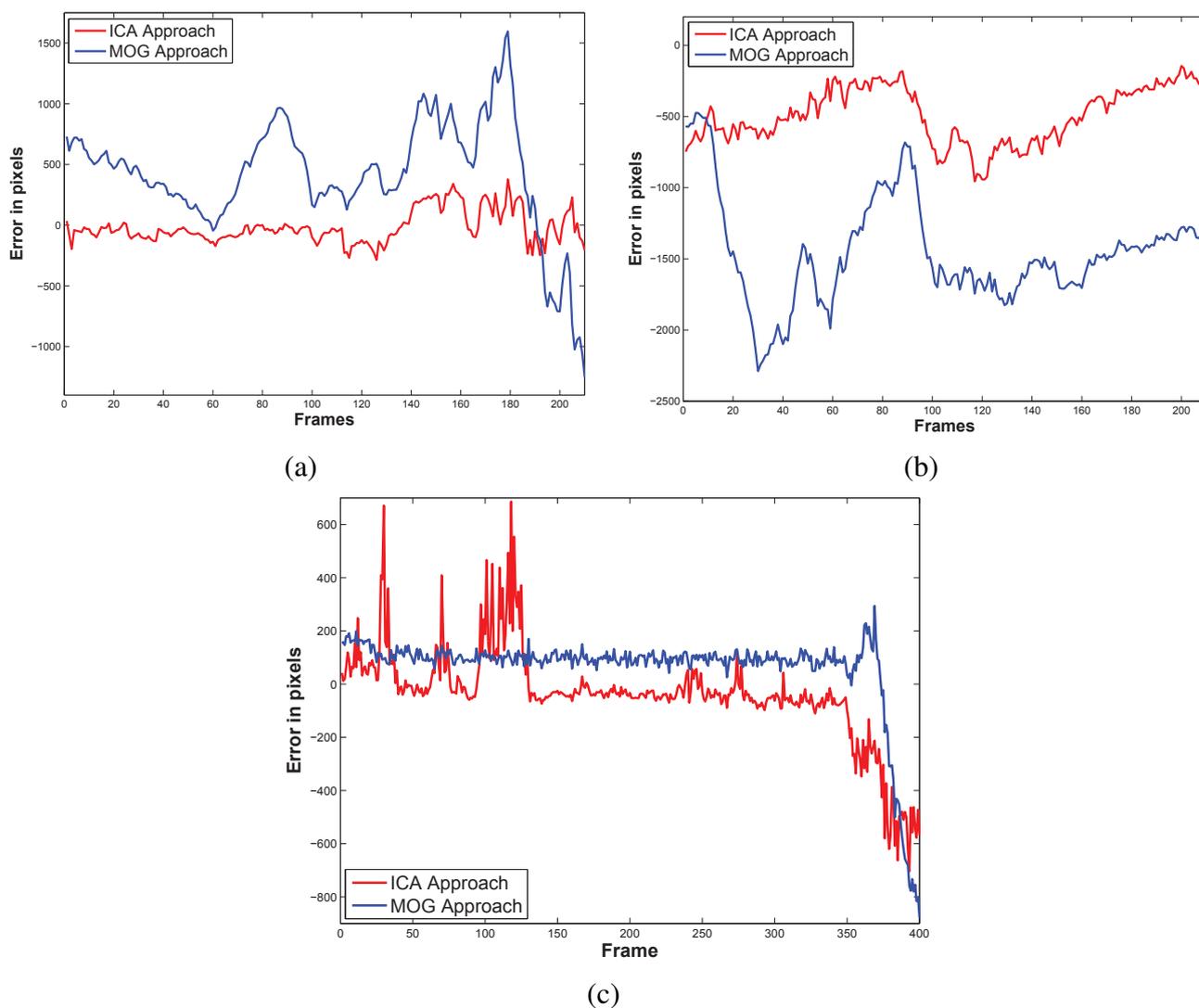
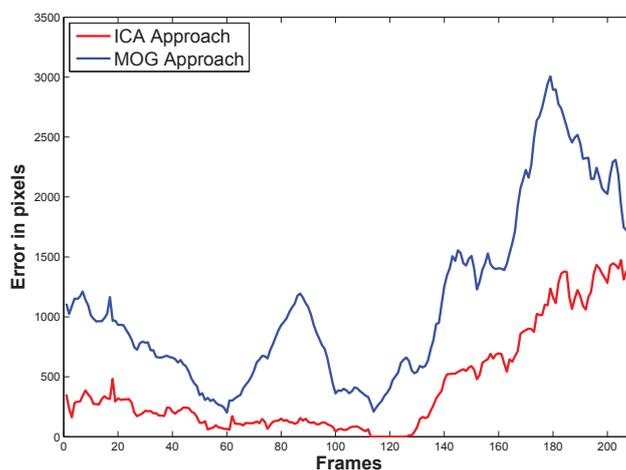
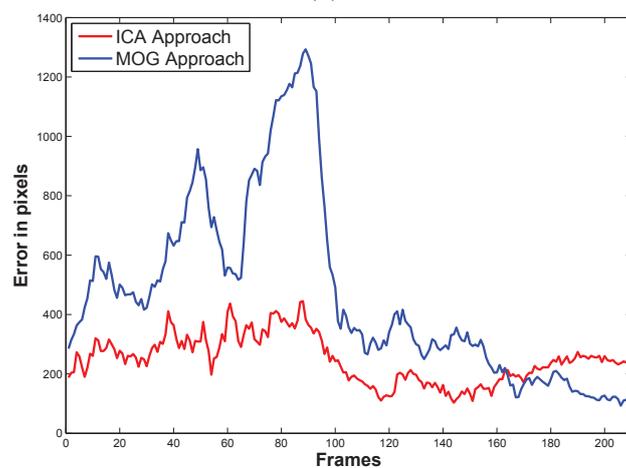


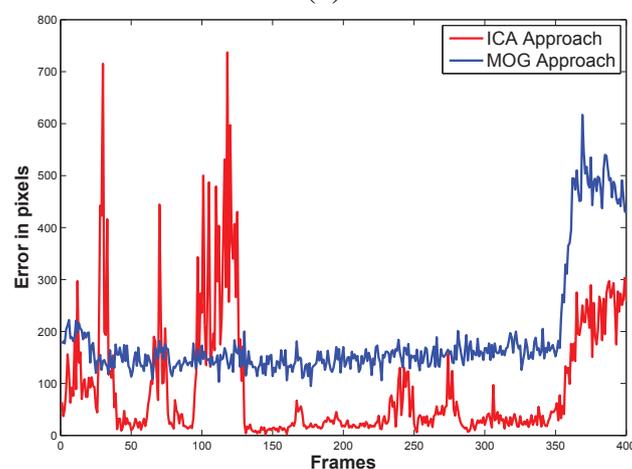
Figure 5. False-positive errors resulted after comparing ground truth against our approach and the MOG approach. In (a), the first train station scenario, where the MOG approach is more sensitive to introduce noise; on the other hand, the proposal adds less error of motion. In (b) the second scenario of the train station at a different perspective, the behavior is similar, except for the end, our proposal has a significant major error bias. In outdoor scenarios (c), the MOG is more sensitive to luminance variations than our proposal.



(a)

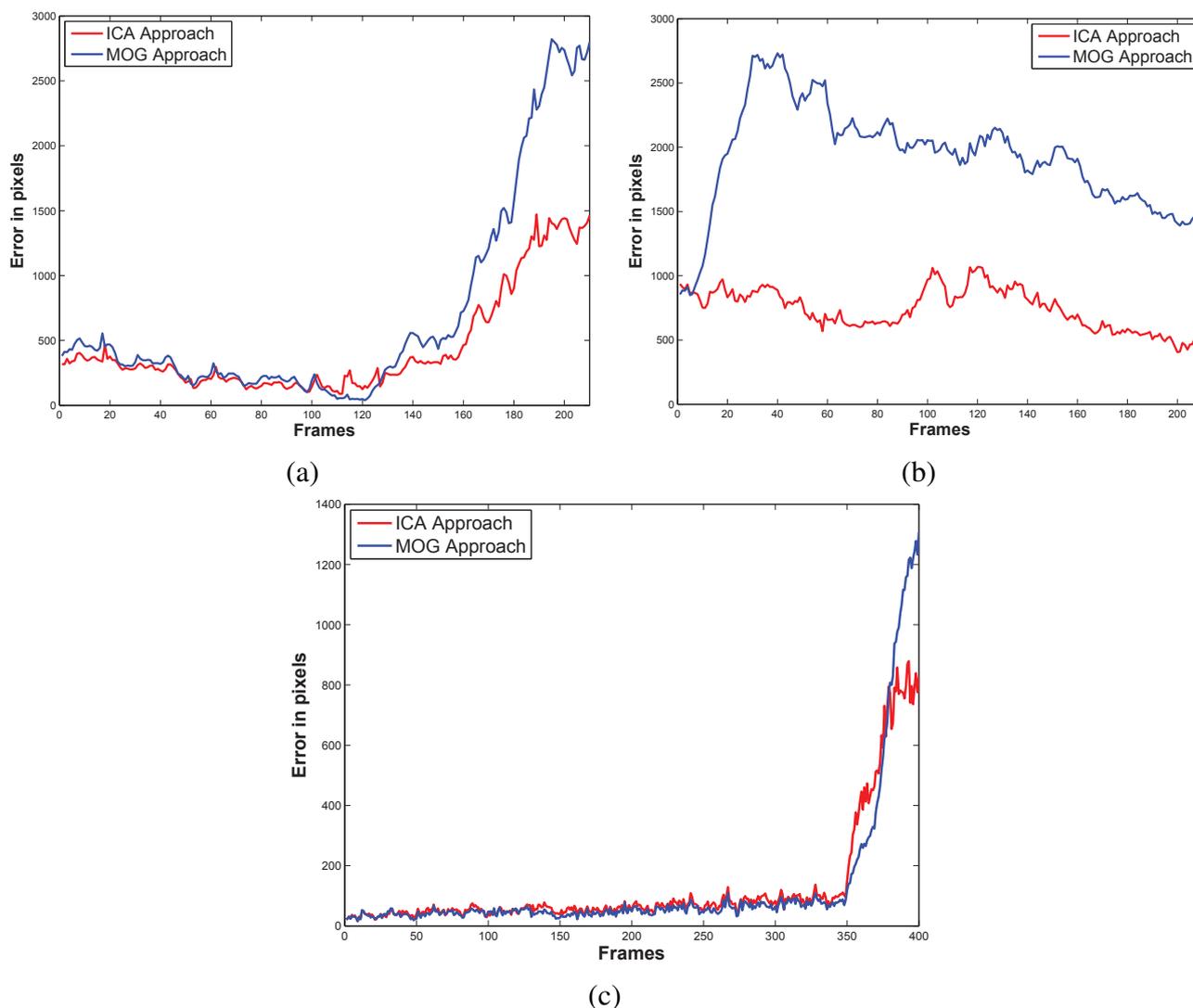


(b)



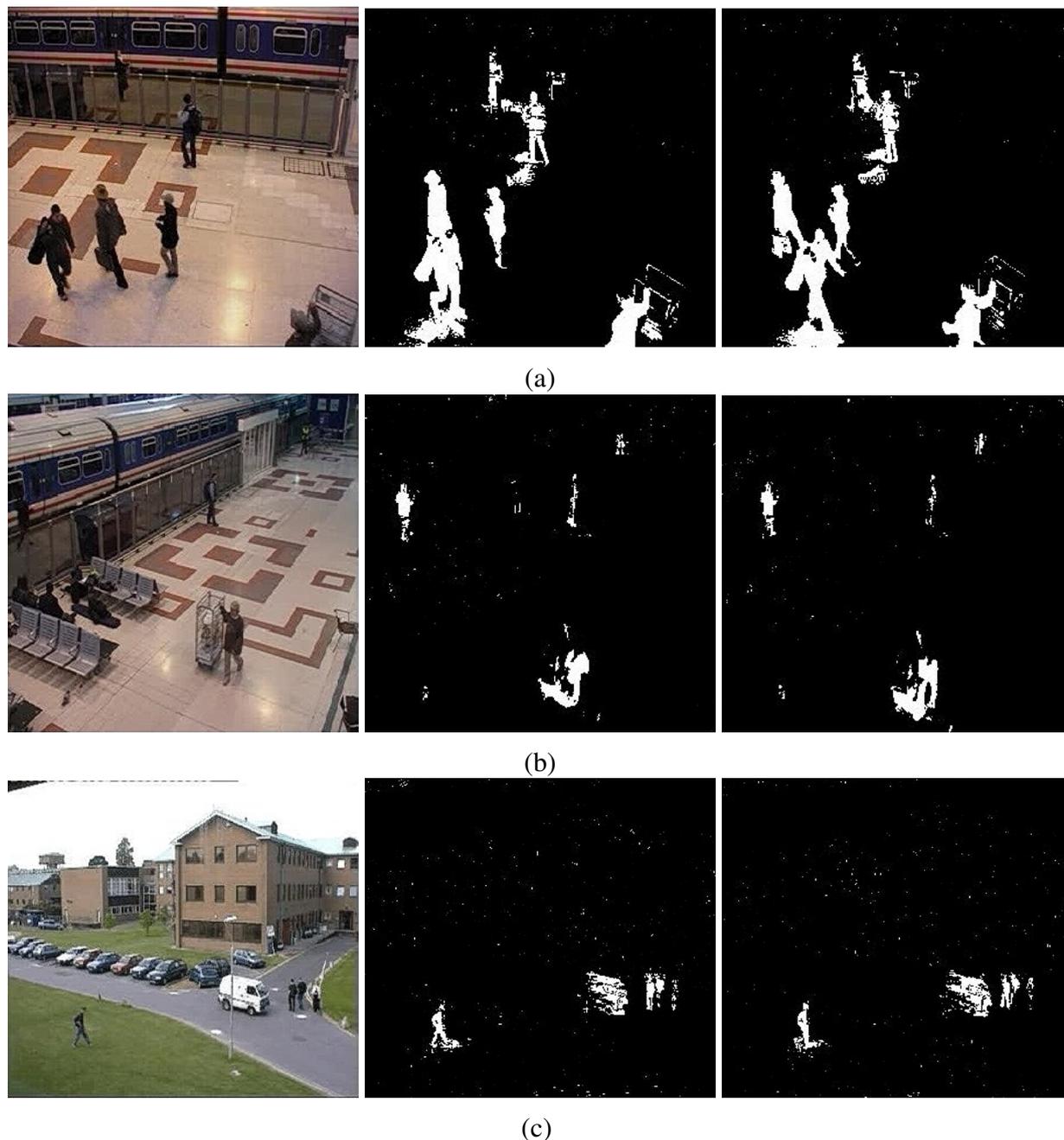
(c)

Figure 6. False-negative errors resulted after comparing ground truth against the proposal and the MOG approach. In the first train station sequence, (a) both errors are similar except for the end of sequence, where our approach has better accuracy for segmenting moving objects. In second train station sequence (b), the MOG approach is more sensitive to people shadows and reflection. Instead, the proposal always has better accuracy for detecting object moving. Finally, in (c) both measures are similar, except at the end of sequence, where our approach offers better accuracy than the MOG approach.



Additionally, the proposal is tested with different image sequences from distinct scenarios. The sequences of images correspond to outdoor scenarios, where the luminance conditions are changeable. The sequences of images were taken from gardens with people walking down and at a vehicular intersection. Figure 8 shows frames belonging to different outdoor sequences. The results show that the proposal is capable to identify foreground objects efficiently. Figure 8a,b shows people walking down in outdoor gardens. Both experiments use the median operator as estimator $E(\Upsilon)$ and $\frac{1}{a_1} \log \cosh a_1 u$ with $a_1 = 1$ to estimate neg-entropy. The images present few noise effects caused by the reflectance and luminance conditions of different objects and materials. In Figure 8c, the object motion corresponds to vehicles at intersections. The scene is affected by vehicle reflections and cloud shadows. The foreground

Figure 7. Foreground detection from images sequences of the PETS database [17, 18] using the MOG approach[3]. The sequences in (a) and (b) correspond to cameras monitoring a train stop; the sequence in (c) corresponds to a car parking. The objects detected are not well defined, and in several frames, some objects are discarded (especially in thin object structures). The motion detection is sensitive to shadows and images are pruned with noise.



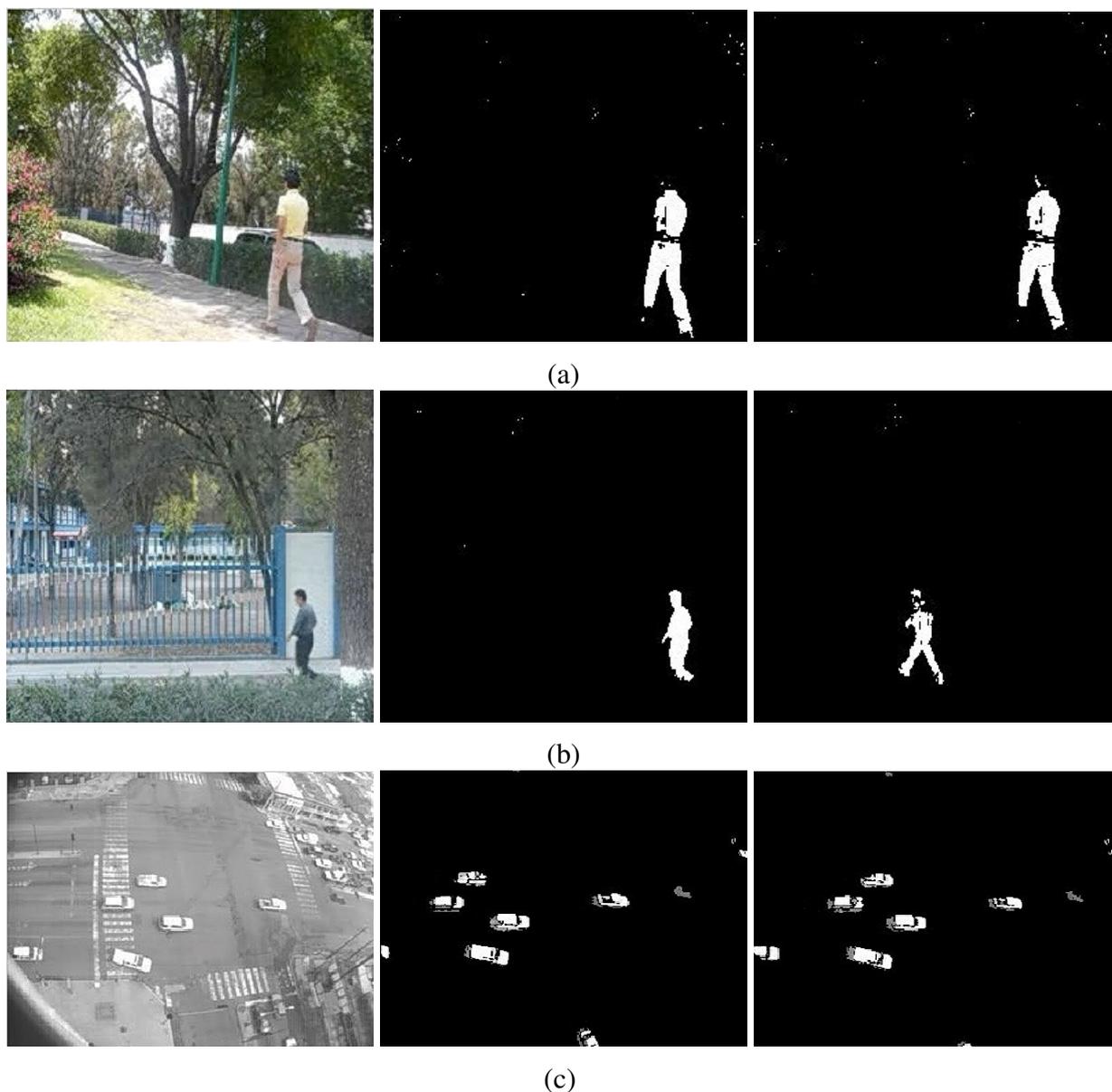
zones are detected efficiently with a weight-average estimator for $E(\Upsilon)$ and a Gaussian for neg-entropy. Finally, in Figure 9, it is presented a scenario with sudden luminance changes. Figure 9a shows some frames used to illustrate the environment conditions. The floor and the wall are affected by soft shadows caused by a woman walking down and the sudden light turned on. Figure 9b shows the results with the MOG approach. The soft shadows, reflections and sudden luminance changes affect negatively the

foreground detection. However, it is appreciated that the approach is capable to discard the majority of the soft shadows and reflections. As an example, when light is turned on, foreground detection is performed adequately.

The ICA approach can be considered as an extension of the PCA approach (see Appendix A), in this sense, an approximation of the motion detection process could be performed using PCA approach. The results are similar when the second order independence is equivalent; in other cases, they are different. This is, when in the PCA approach, the principal components are both, independent and uncorrelated, the PCA is equivalent to the ICA. In the practice, both are equivalent when the pixels suffer global effects. But, when scenario have several luminance sources, there are some considerable differences. To illustrate it, Figure 10 shows the differences of I^* and the second principal component estimated with the PCA approach, using the the sequence of the intersection. The images are in pseudo-color where dark zones represent small differences, and red color, considerable differences. The main differences are detected in zones with shadows and reflecting materials. The proposal identifies the objects better as the local distribution of pixels values must be independent and uncorrelated of the estimation $E(Y)$; in contrary, the PCA approach is enough that data are independent. In the practice, the differences depend of the optimization process to estimate W and the approximation of neg-entropy used. But, the results are always better in the proposal, given that it adds the constraint of uncorrelation that the PCA does not consider.

As additional results, the binary motion map $B(\mathbf{x})$ is improved with a opening by reconstruction filter. This filter helps out to clarify the motion zones and eliminates noise classification regions. Figure 11 shows some frames acquired from a public garden, where luminance conditions are changeable. The motion detection is performed, but it is affected with noise. The morphological filter eliminates those regions low-connected. The resulting image offers a better definition of the motion objects. However the filters must be used carefully, because they require more computational resources and affects the frame rate specially in real time systems.

Figure 8. Foreground detection from (a) a camera monitoring a public garden, (b) a camera monitoring a people lane, (c) a camera monitoring an intersection. The first two images show several leaves moving caused by wind and the scene is continuously changing. The third one, shows cloud shadows causing changes of luminance conditions.



The proposal method works efficiently to detect foreground objects when data information that represents the foreground is separable from the background. The proposal method offers a numerical approximation, and in practice, it needs to consider some points that could affect the foreground detection. The level of accuracy depends of the convergence criterion of the algorithm showed in §3.3, the number of components that are considered as noise data in Σ^* , and the amount of information to estimate each independent component; consequently, the small foreground zones are difficult to be detected. As an example, we can see in Figures 3 and 6 where a person walking down is not completely detected.

Figure 9. Foreground detection at a scene with sudden luminance change. In (a), we present a room, where light is turning on, the floor shows reflections and the wall shows soft shadows. (b) When the MOG approach is applied, the reflections and the shadows are detected as moving objects. (c) Instead, the proposal is capable to detect better the human silhouette without the introduction of moving noise, even the light condition has changed.

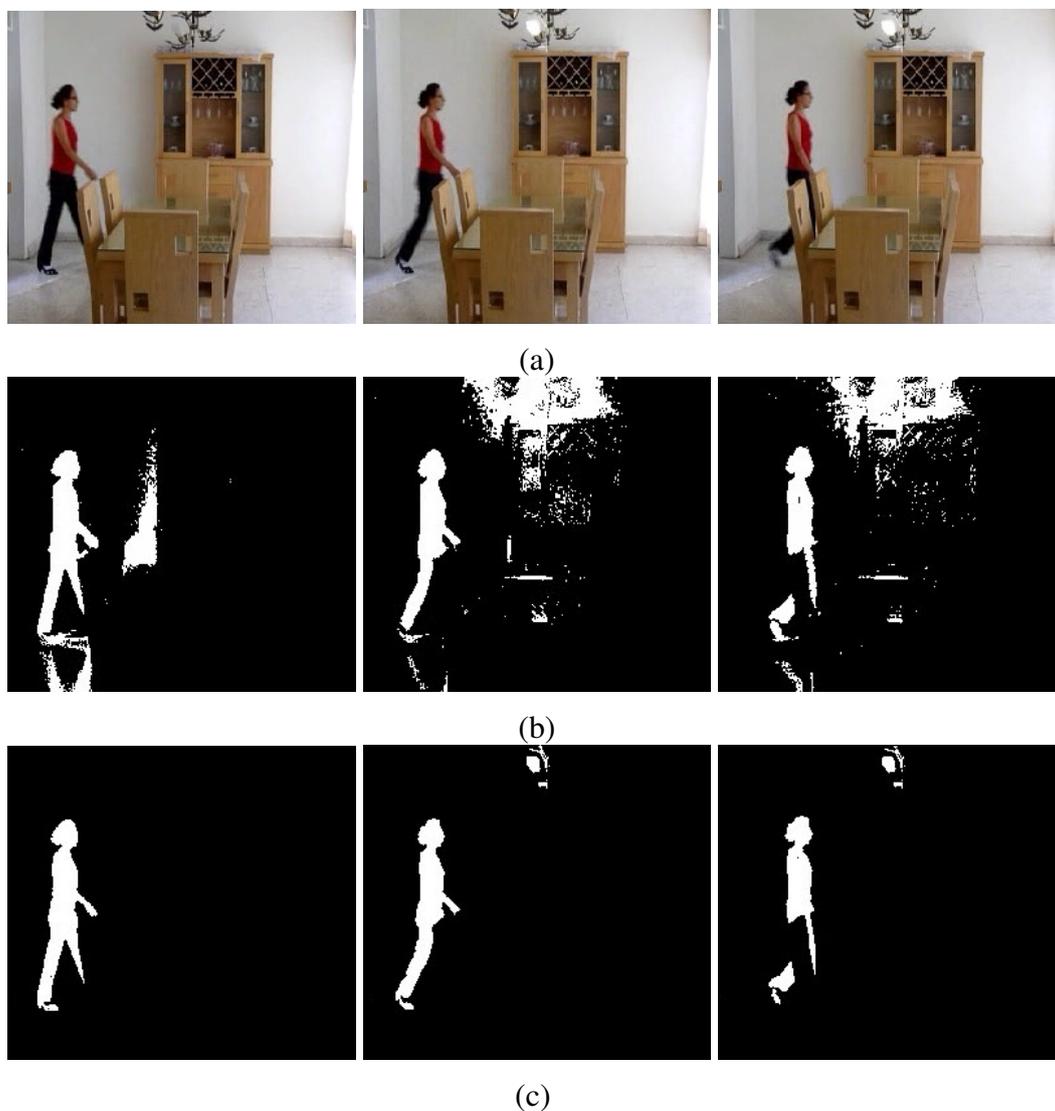
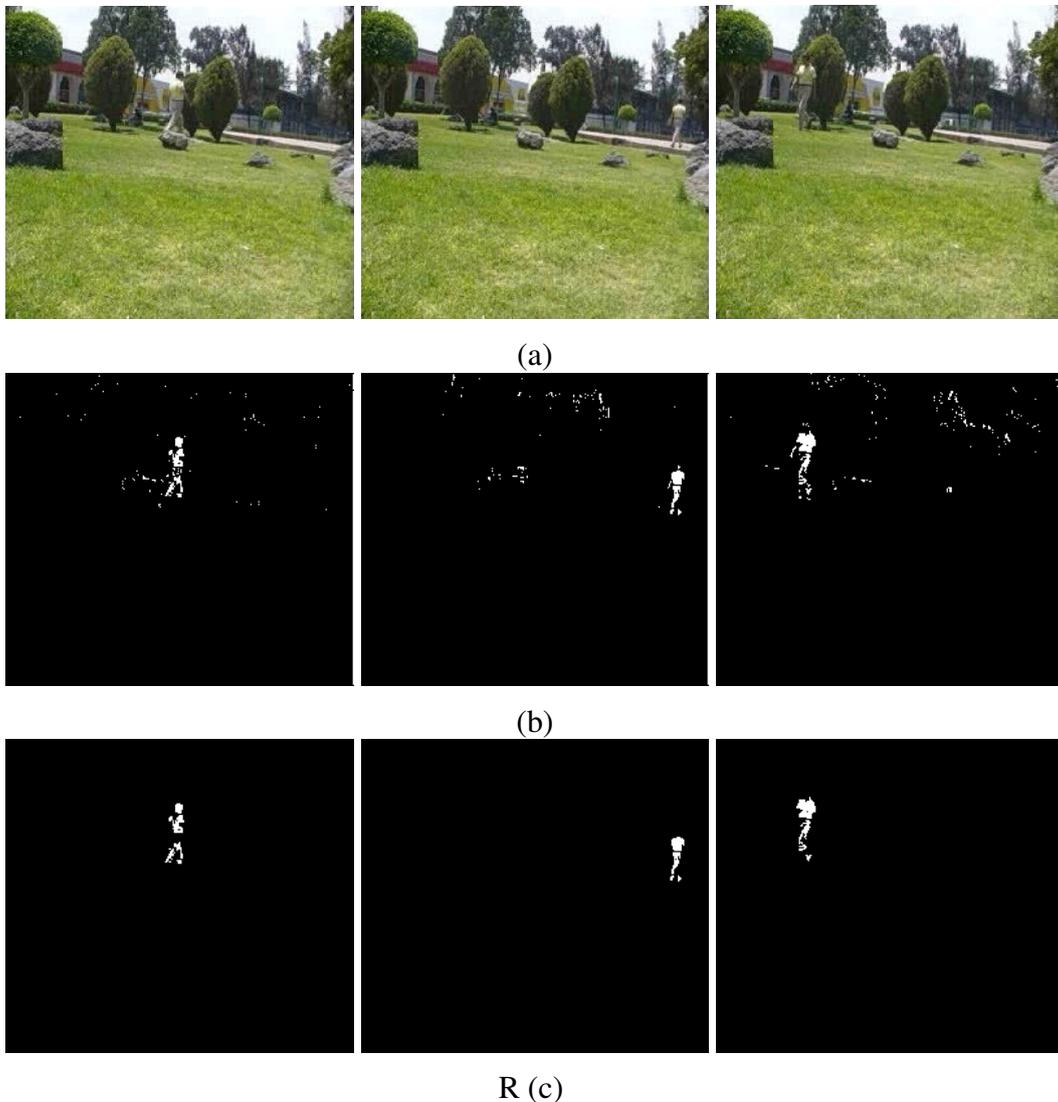


Figure 10. Images differences amongst the components estimated with the proposal and the PCA estimation. The main differences correspond to the areas affected by several lights, as the roof of the vehicles and the buildings, and, in minor sense, the global lights effects over the scenario.



Figure 11. The enhancement of binary motion maps. The motion objects are better defined, and the noise effects are discarded. In (a) the original frames; in (b) the estimation of $B(\mathbf{x})$, and in (c) the enhancement map $\tilde{B}(\mathbf{x})$.



6. Conclusions

In this work, we presented a background subtraction approach based on an independent component analysis. This approach exploits the separability and the non-Gaussianity to estimate each of the pdfs. The motion is efficiently detected when raw data are separable. The results show a robust approach at several scenarios. The moving objects are detected with high resolution even when the scenario presents luminance changes or shadows. The implementation issues have been discussed and this approach can be implemented in a real time monitoring system. Finally, this approach has a better accuracy than the MOG approach in the tested scenarios and is superior in scenarios with changing environmental conditions.

Acknowledgements

We give thanks to the CIDESI Research Center for the support and material provided to develop this research.

References

1. Toyama, K.; Krumm, J.; Brumitt, B.; Meyers, B. Wallflower: Principles and Practice of Background Maintenance. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Ft. Collins, CO, USA, June 23–25, 1999; Volume 1, pp. 255–261.
2. Horprasert, T.; Harwood, D.; Davis, L.S. A Robust Background Subtraction and Shadow Detection. In *Proceedings of Asian Conference on Computer Vision*, Taipei, Taiwan, February 27–March 3, 2000.
3. Stauffer, C.; Grimson, W. Adaptive Background Mixture Models for Real-Time Tracking. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Ft. Collins, CO, USA, June 23–25, 1999; Volume 2, pp. 252–259.
4. Cheung, S.C.S.; Kamath, C. Robust Background Subtraction with Foreground Validation for Urban Traffic Video. *EURASIP J. Appl. Signal Process.* **2005**, *2005*, 2330–2340.
5. KaewTraKulPong, P.; Bowden, R. An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection. In *Proceedings of the 2nd European Workshop on Advanced Video Based Surveillance Systems*, London, UK, September 4, 2001; pp. 149–158.
6. Elgammal, A.; Harwood, D.; Davis, L.S. Non-parametric Model for Background Subtraction. In *Proceedings of IEEE International Conference on Computer Vision*, Kerkyra, Corfu, Greece, September 20–25, 1999.
7. Oliver, N.M.; Rosario, B.; Pentland, A.P. A Bayesian Computer Vision System for Modeling Human Interactions. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 831–843.
8. Shaw, P.J.A. *Multivariate Statistics for the Environmental Sciences*; Hodder-Arnold: London, UK, 2003.
9. Rymel, J.; Renno, J.; Greenhill, D.; Orwell, J.; Jones, G. Adaptive Eigen-Backgrounds for Object Detection. In *Proceedings of IEEE International Conference on Image Processing*, Singapore, October 24–27, 2004; Volume 3, pp. 1847–1850.
10. Han, B.; Comaniciu, D.; Davis, L. Sequential Kernel Density Approximation through Mode Propagation: Applications to Background Modeling. In *Proceedings of Asian Conference on Computer Vision*, Jeju Island, Korea, January 2004.
11. Tang, Z.; Miao, Z. Fast Background Subtraction and Shadow Elimination Using Improved Gaussian Mixture Model. In *Proceedings of the 6th IEEE International Workshop on Haptic, Audio and Visual Environments and Games*, Ottawa, ON, Canada, October 12–14, 2007; pp. 38–41.
12. Zhou, D.; Zhang, H.; Ray, N. Texture based Background Subtraction. In *Proceedings of International Conference on Information and Automation*, Zhangjiajie, Hunan, China, June 20–23, 2008; pp. 601–605.
13. Tsai, D.M.; Lai, S.C. Independent Component Analysis-Based Background Subtraction for Indoor Surveillance. *IEEE Trans. Image Process.* **2009**, *18*, 158–167.

14. Hyvarinen, A. Survey on Independent Component Analysis. *Neural Comput. Surv.* **1999**, *2*, 94–128.
15. Comon, P. Independent Component Analysis—A New Concept? *Signal Process.* **1994**, *36*, 287–314.
16. Suli, E.; Mayers, D. *An Introduction to Numerical Analysis*; Cambridge University Press: Cambridge, UK, 2003.
17. The University of Reading. Performance Evaluation of Tracking and Surveillance (PETS). Internet, <ftp://ftp.pets.rdg.ac.uk/pub/PETS2001/>, 2001. The University of Reading, UK.
18. Thirde, D.; Ferryman, J.; Crowley, J.L. Performance Evaluation of Tracking and Surveillance (PETS). Available online: <http://pets2006.net/> (accessed on 5 February 2010).
19. Kalman, R. A New Approach to Linear Filtering and Prediction Problems. *J. Basic Eng.* **1960**, *82*, 35–45.
20. Dempster, A.; Laird, N.; Rubin, D. Maximum Likelihood from Incomplete Data via the EM Algorithm. *JR Statist. Soc.* **1977**, *39*, 1–38.
21. Hyvainen, A. Fast and Robust Fixed-Point Algorithms for Independent Component Analysis. *IEEE Trans. Neural Networks* **1999**, *10*, 626–634.
22. Harry, L.; Papadimitrou, C. *Elements of the Theory of Computation*; Prentice Hall: Upper Saddle River, NJ, USA, 1988.
23. Russell, S.; Norvig, P. *Artificial Intelligence: A Modern Approach*; Prentice Hall: Upper Saddle River, NJ, USA, 2003.
24. Salembier, P.; Serra, J. Flat Zones Filtering, Connected Operators and Filters by Reconstruction. *IEEE Trans. Image Process.* **1995**, *4*, 1153–1160.
25. Bao, P.; Zhang, L. Noise Reduction for Multi-scale Resonance Images via Adaptive Multiscale Products Thresholding. *IEEE Trans. Med. Imaging* **2003**, *2*, 1089–1099.
26. Pless, R.; Larson, J.; Siebers, S.; Westover, B. Evaluation of Local Models of Dynamic Backgrounds. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Madison, WI, USA, June 16–22, 2003; Volume 2, pp. 73–78.
27. Herrero, S.; Bescós, J. Background Subtraction Techniques: Systematic Evaluation and Comparative Analysis. In *Advanced Concepts for Intelligent Vision Systems*; Springer Verlag: Berlin, Germany, 2009.
28. Papoulis, A. *Probability, Random Variables, and Stochastic Processes*; McGraw-Hill: New York, NY, USA, 1991.
29. Duda, R.O.; Hart, P.E.; Stork, D.G. *Pattern Classification*; Wiley-Interscience: Malden, MA, USA, 2000.

A Motion Detection and PCA approach

The PCA approach is used as a subspace projection technique. In PCA, the basis vectors are obtained by solving the algebraic eigenvalue system $R^T(\mathbf{X}\mathbf{X}^T)R = \Sigma$, where X is the centered data, R is a matrix of eigenvectors, and Σ is the corresponding diagonal matrix of eigenvalues. The projection of the data $\mathbf{C}_n = R_n^T \mathbf{X}$, from the original p dimensional space to a subspace spanned by n principal eigenvectors is optimal under the mean squared error; *i.e.*, the projection of \mathbf{C}_n back into the p dimensional space, has a

minimum reconstruction error. In fact, if n is large enough to include all the eigenvectors with non-zero eigenvalues, the projection is lossless.

Thus, the goal in the PCA approach is to minimize the projection error from compressed data, the goal of the ICA approach is to minimize the statistical dependence between the basis vectors; *i.e.*, this can be written as $W\mathbf{X}^T = \mathbf{U}$, where the ICA approach searches for a linear transformation W that minimizes the statistical dependence between the rows of \mathbf{U} , given a training set \mathbf{X} . Unlike PCA, the basis vectors in ICA are neither orthogonal nor ranked in order. Also, there is no closed form expression to find W . This work is based on the FastICA approach [21], which use the neg-entropy to estimate W , given that it is invariant to linear transformation [15]. This is, the estimation of W minimizes the mutual information, resulting roughly equivalent to find the directions in which the neg-entropy is maximized. This formulation shows explicitly the connection between the ICA and projection pursuit. In fact, finding a single direction that maximizes neg-entropy is a form of projection pursuit, and could also be interpreted as estimation of a single independent component. Thus ICA is an extension of PCA approach with the adds of the imposing independence up to the second order and, the definition of directions that are orthogonal as discussed in [15].

The Motion Detection process uses the middle-significant eigenvalues of the mixing matrix W^{-1} ; *i.e.*, there is a dimensionality reduction. However, the reduction is performed in the unmixed data \mathbf{U} . Then, the dimensionality reduction on this matrix could result equivalent to PCA, when the second order independence are equivalent to the obtained with the PCA estimation.

© 2010 by the authors; licensee MDPI, Basel, Switzerland. This article is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution license <http://creativecommons.org/licenses/by/3.0/>.