

Research Article

Robust Real-Time Background Subtraction Based on Local Neighborhood Patterns

Ariel Amato, Mikhail G. Mozerov, F. Xavier Roca, and Jordi González

Computer Vision Center (CVC), Universitat Autònoma de Barcelona, Campus UAB Edifici O, 08193 Bellaterra, Spain

Correspondence should be addressed to Mikhail G. Mozerov, mozerov@cvc.uab.es

Received 1 December 2009; Accepted 21 June 2010

Academic Editor: Yingzi Du

Copyright © 2010 Ariel Amato et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper describes an efficient background subtraction technique for detecting moving objects. The proposed approach is able to overcome difficulties like illumination changes and moving shadows. Our method introduces two discriminative features based on angular and modular patterns, which are formed by similarity measurement between two sets of RGB color vectors: one belonging to the background image and the other to the current image. We show how these patterns are used to improve foreground detection in the presence of moving shadows and in the case when there are strong similarities in color between background and foreground pixels. Experimental results over a collection of public and own datasets of real image sequences demonstrate that the proposed technique achieves a superior performance compared with state-of-the-art methods. Furthermore, both the low computational and space complexities make the presented algorithm feasible for real-time applications.

1. Introduction

Moving object detection is a crucial part of automatic video surveillance systems. One of the most common and effective approach to localize moving objects is background subtraction, in which a model of the static scene background is subtracted from each frame of a video sequence. This technique has been actively investigated and applied by many researchers during the last years [1–3]. The task of moving object detection is strongly hindered by several factors such as shadows cast by moving object, illuminations changes, and camouflage. In particular, cast shadows are the areas projected on a surface because objects are occluding partially or totally direct light sources. Obviously, an area affected by cast shadow experiences a change of illumination. Therefore in this case the background subtraction algorithm can misclassify background as foreground [4, 5]. Camouflage occurs when there is a strong similarity in color between background and foreground; so foreground pixels are classified as background. Broadly speaking, these issues rise problems such as shape distortion, object merging, and even object losses. Thus a robust and accurate algorithm to segment moving object is highly desirable.

In this paper, we present an adaptive background model, which is formed by temporal and spatial components. These

components are basically computed by measuring the angle and the Euclidean distance between two sets of color vectors. We will show how these components are combined to improve the robustness and the discriminative sensitivity of the background subtraction algorithm in the presence of (i) moving shadows and (ii) strong similarities in color between background and foreground pixels. Another important advantage of our algorithm is its low computational complexity and its low space complexity that makes it feasible for real-time applications.

The rest of the paper is organized as follows. Section 2 introduces a brief literature review. Section 3 presents our method. In Section 4 experimental results are discussed. Concluding remarks are available in Section 5.

2. Related Work

Many publications are devoted to the background subtraction technique [1–3]. However in this section we consider only the papers that are directly related to our work.

Haritaoglu et al. state that in W4 [6] the background is modeled by representing each pixel by three values: its minimum and maximum intensity values and the maximum intensity differences between consecutive frames observed

during this training period. Pixels are classified as foreground if the differences between the current value and the minimum and maximum values are greater than the values of the maximal interframe difference. However, this approach is rather sensitive to shadows and lighting changes, since the only illumination intensity cue is used and the memory resource to implement this algorithm is extremely high.

Horprasert et al. [7] implement a statistical color background algorithm, which use color chrominance and brightness distortion. The background model is built using four values: the mean, the standard deviation, the variation of the brightness, and chrominance distortion. However, this approach usually fails for low and high intensities.

Kim et al. [8] use a similar approach as [7], but they obtain more robust motion segmentation in the presents of the illumination and scene changes using background model with codebooks. The codebooks idea gives the possibility to learn more about the model in the training period. The authors propose to cope with the unstable information of the dark pixels, but still they have some problems in the low- and the high-intensity regions. Furthermore, the space complexity of their algorithm is high.

Stauffer and Grimson [9] address the low- and the high-intensity regions problem by using a mixture of Gaussians to build a background color model for every pixel. Pixels from the current frame are checked against the background model by comparing them with every Gaussian in the model until a matching Gaussian is found. If so, the mean and variance of the matched Gaussian are updated; otherwise a new Gaussian with the mean equal to the current pixel color and some initial variance is introduced into the mixture.

McKenna et al. [10] assume that cast shadows result in significant change in intensity without much change in chromaticity. Pixel chromaticity is modeled using its mean and variance and the first-order gradient of each background pixel modeled using gradient means and magnitude variance. Moving shadows are then classified as background if the chromaticity or gradient information supports their classification.

Cucchiara et al. [11] use a model in Hue-Saturation-Value (HSV) and stress their approach in shadow suppression. The idea is that shadows change the hue component slightly and decrease the saturation component significantly. In the HSV color space a more realistic noise model can be done. However, this approach also has drawbacks. The similarity measured in the nonlinear HSV color space usually generates ambiguity at gray levels. Furthermore threshold handling is the major limitation of this approach.

3. Proposed Algorithm

A simple and common background subtraction procedure involves subtraction of each new image from a static model of the scene. As a result a binary mask with two labels (foreground and background) is formed for each pixel in the image plane. Broadly speaking, this technique can be

separated in two stages, one dealing with the scene modeling and another with the motion detection process. The scene modeling stage represents a crucial part in the background subtraction technique [12–17].

Usually a simple unimodal approach uses statistical parameters such as mean and standard deviation values, for example, [7, 8, 10], and so forth. Such statistical parameters are obtained during a training period and then these are dynamically updated. In the background modeling process the statistical values depend on both the low- and high-frequency changes of the camera signal. If the standard deviations of the low- and high-frequency components of the signal are comparable, methods based on such statistical parameters exhibit robust discriminability. When the standard deviation of the high-frequency change is significantly less than the low-frequency change, then the background model can be improved to make the discriminative sensitivity much higher. Since a considerable change in the low-frequency domain is produced for the majority of real video sequences, we propose to build a model that is insensitive to low-frequency changes. The main idea is to estimate only the high-frequency change per each pixel value as one interframe interval. The general background model in this case can be explained as the subtraction between the current frame and the previous frame, which suppose to be the background image. Two values for each pixel in the image are computed to model background changes during the training period: the maximum difference in angular and Euclidean distances between the color vectors of the consecutive image frames. The angular difference is used because it can be considered as photometric invariant of color measurement and in turn as significant cues to detect moving shadows.

Often pixelwise comparison is not enough to distinguish background from foreground and in our classification process we further analyze the neighborhood of each pixel position. In the next section we give a formal definition of the proposed similarity measurements.

3.1. Background Scene Modeling

3.1.1. Similarity Measurements. Four similarity measurements are used to compare a background image with a current frame.

- (i) *Angular similarity measurement* $\Delta\theta$ between two color vectors $\mathbf{p}(\mathbf{x})$ and $\mathbf{q}(\mathbf{x})$ at position \mathbf{x} in the RGB color space is defined as follows:

$$\Delta\theta(\mathbf{p}(\mathbf{x}), \mathbf{q}(\mathbf{x})) = \text{Cos}^{-1} \left(\frac{\mathbf{p}(\mathbf{x}) \cdot \mathbf{q}(\mathbf{x})}{|\mathbf{p}(\mathbf{x})| |\mathbf{q}(\mathbf{x})|} \right). \quad (1)$$

- (ii) *Euclidean distance similarity measurement* ΔI between two color vectors $\mathbf{p}(\mathbf{x})$ and $\mathbf{q}(\mathbf{x})$ in the RGB color space is defined as follows:

$$\Delta I(\mathbf{p}(\mathbf{x}), \mathbf{q}(\mathbf{x})) = |\mathbf{p}(\mathbf{x}) - \mathbf{q}(\mathbf{x})|. \quad (2)$$

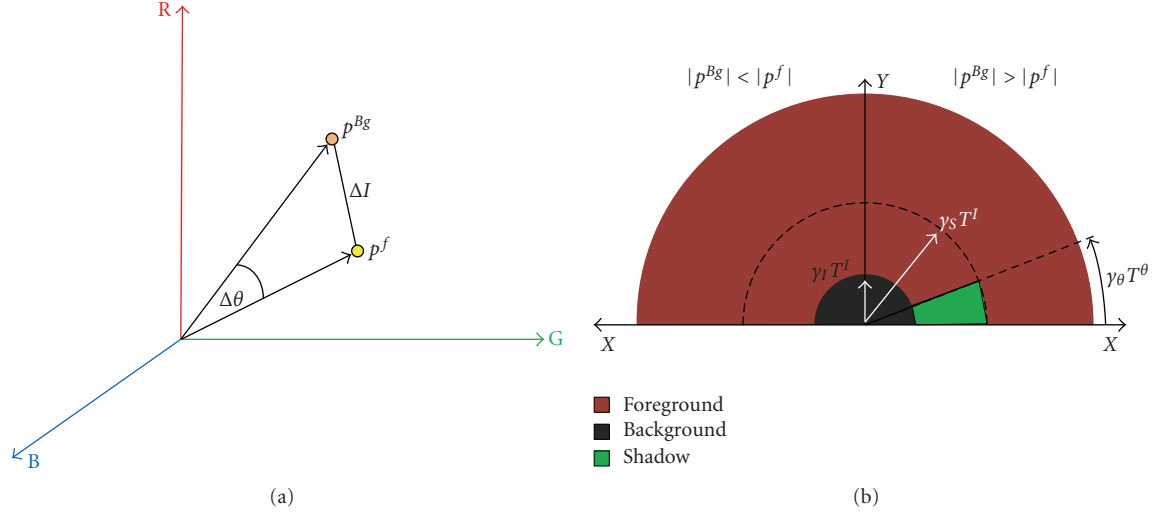


FIGURE 1: (a) Angle and magnitude difference between two color vector in RGB space. (b) Difference in angle and magnitude in 2D ‘‘polar difference space.’’ The axes are computed as $x = \Delta I \cdot \cos(\Delta\theta)$ and $y = \Delta I \cdot \sin(\Delta\theta)$.

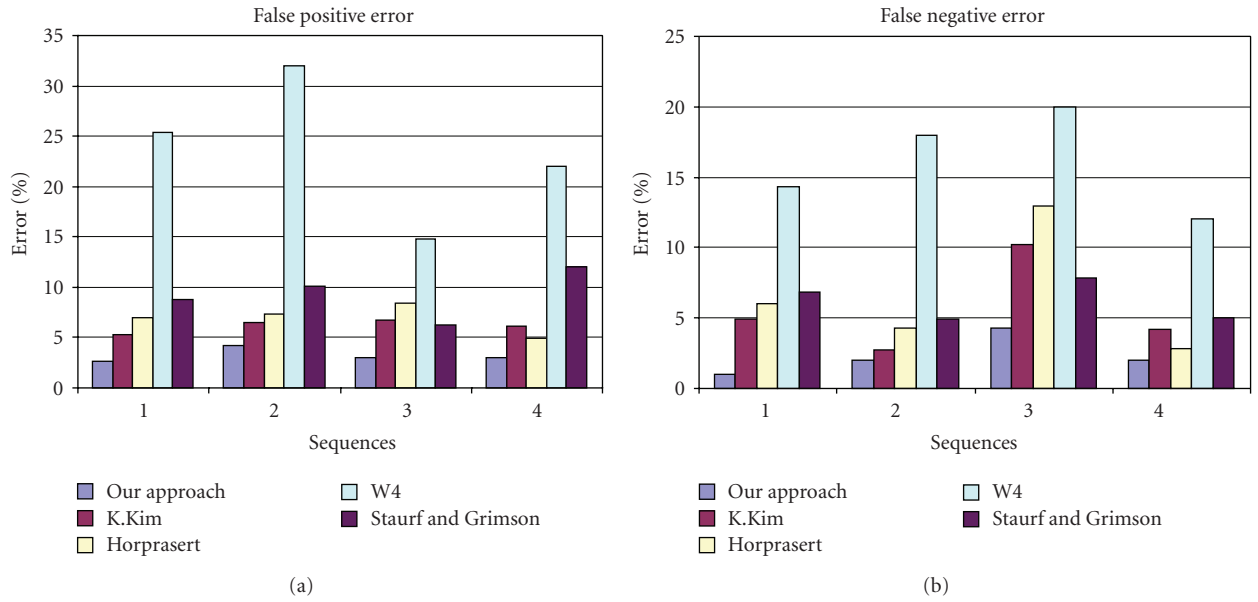


FIGURE 2: Segmentation errors. (a) FPE and (b) FNE.

For each of the described similarity measurements a threshold function is associated:

$$T\theta(\Delta\theta, \theta^T) = \begin{cases} 1, & \text{if } \Delta\theta > \theta^T, \\ 0, & \text{otherwise,} \end{cases} \quad (3)$$

$$TI(\Delta I, I^T) = \begin{cases} 1, & \text{if } |\Delta I| > I^T, \\ 0, & \text{otherwise,} \end{cases}$$

where θ^T and I^T are intrinsic parameters of the threshold functions of the similarity measurements.

To describe a neighbourhood similarity measurement let us first characterize the index vector $\mathbf{x} = (\mathbf{n}, \mathbf{m})^t \in \Omega = \{0, 1, \dots, n, \dots, N; 0, 1, \dots, m, \dots, M\}$, which define the position of a pixel in the image. Also we need to name the neighbourhood radius vector $\mathbf{w} = (\mathbf{i}, \mathbf{j})^t \in \mathbf{W} = \{-W, \dots, 0, 1, \dots, i, \dots, W; -W, \dots, 0, 1, \dots, j, \dots, W\}$, which define the positions of pixels that belong to the neighbourhood relative to any current pixel. Indeed, the domain \mathbf{W} is just a square window around a chosen pixel.

(iii) *Angular neighborhood similarity measurement* $\eta\theta$ between two sets of color vectors in the RGB color

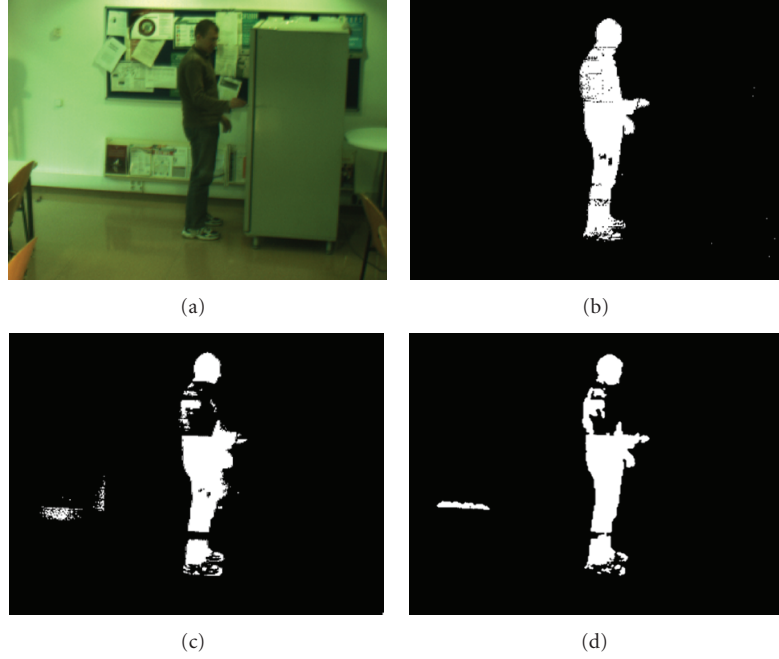


FIGURE 3: (a) Original image, segmentation result of (b) our method, (c) Stauffer method, and (d) K. Kim method.

space $\mathbf{p}(\mathbf{x} + \mathbf{w})$ and $\mathbf{q}(\mathbf{x} + \mathbf{w})$ ($\mathbf{w} \in \mathbf{W}$) can be written as

$$\eta\theta(\vartheta, \theta^T) = \sum_{\mathbf{w} \in \mathbf{W}} T\theta(\Delta\theta(\vartheta), \theta^T), \quad (4)$$

where $T\theta$, θ^T , and $\Delta\theta$ are defined in (3) and (1), respectively, and ϑ is $(\mathbf{p}(\mathbf{x} + \mathbf{w}), \mathbf{q}(\mathbf{x} + \mathbf{w}))$.

- (iv) *Euclidean distance neighborhood similarity measurement* μI between two sets of color vectors in the RGB color space $\mathbf{p}(\mathbf{x} + \mathbf{w})$ and $\mathbf{q}(\mathbf{x} + \mathbf{w})$ ($\mathbf{w} \in \mathbf{W}$) can be written as

$$\mu I(\vartheta, I^T) = \sum_{\mathbf{w} \in \mathbf{W}} TI(\Delta I(\vartheta), I^T), \quad (5)$$

where TI , I^T , and ΔI are defined in (3) and (2), respectively. With each of the neighbourhood similarity measurements we associate a threshold function:

$$\begin{aligned} T\eta\theta(\eta\theta(\vartheta), \eta^T) &= \begin{cases} 1, & \text{if } \eta\theta(\vartheta) > \eta^T, \\ 0, & \text{otherwise,} \end{cases} \\ T\mu I(\mu I(\vartheta), \mu^T) &= \begin{cases} 1, & \text{if } \mu I(\vartheta) > \mu^T, \\ 0, & \text{otherwise,} \end{cases} \end{aligned} \quad (6)$$

where η^T and μ^T are intrinsic parameters of the threshold functions of the neighborhood similarity measurements.

3.1.2. Scene Modeling. Our background model (BG) will be represented with two classes of components, namely,

running components (RCs) and training components (TCs). The RC is a color vector in RGB space and only this component can be updated in running process. The TC is a set of fixed thresholds values obtained during the training.

The background model is represented by

$$\text{BG}(\mathbf{x}) = \left\{ \{\mathbf{p}(\mathbf{x})\}, \{T^\theta(\mathbf{x}), T^I(\mathbf{x}), W\} \right\}, \quad (7)$$

where $T^\theta(\mathbf{x})$ is maxima of the chromaticity variation; $T^I(\mathbf{x})$ is maxima of the intensity variation; W is the half size of the neighbourhood window.

A training process has to be performed to obtain the background parameters defined by (7). This first step consists of estimating the value of the RC and TC during the training period. To initialize our BG we put the $\text{RC} = \{\mathbf{p}^0(\mathbf{x})\}$ as the initial frame. $T^\theta(\mathbf{x})$ and $T^I(\mathbf{x})$ are estimated during the training period by computing the angular difference and the Euclidean distance between the pixel belonging to the previous frame and the pixel belonging to the current frame:

$$\begin{aligned} T^\theta(\mathbf{x}) &= \max_{f \in \{1, 2, \dots, F\}} \left\{ \Delta\theta(\mathbf{p}^{f-1}(\mathbf{x}), \mathbf{p}^f(\mathbf{x})) \right\}, \\ T^I(\mathbf{x}) &= \max_{f \in \{1, 2, \dots, F\}} \left\{ \Delta I(\mathbf{p}^{f-1}(\mathbf{x}), \mathbf{p}^f(\mathbf{x})) \right\}, \end{aligned} \quad (8)$$

where F is the number of frames in the training period.

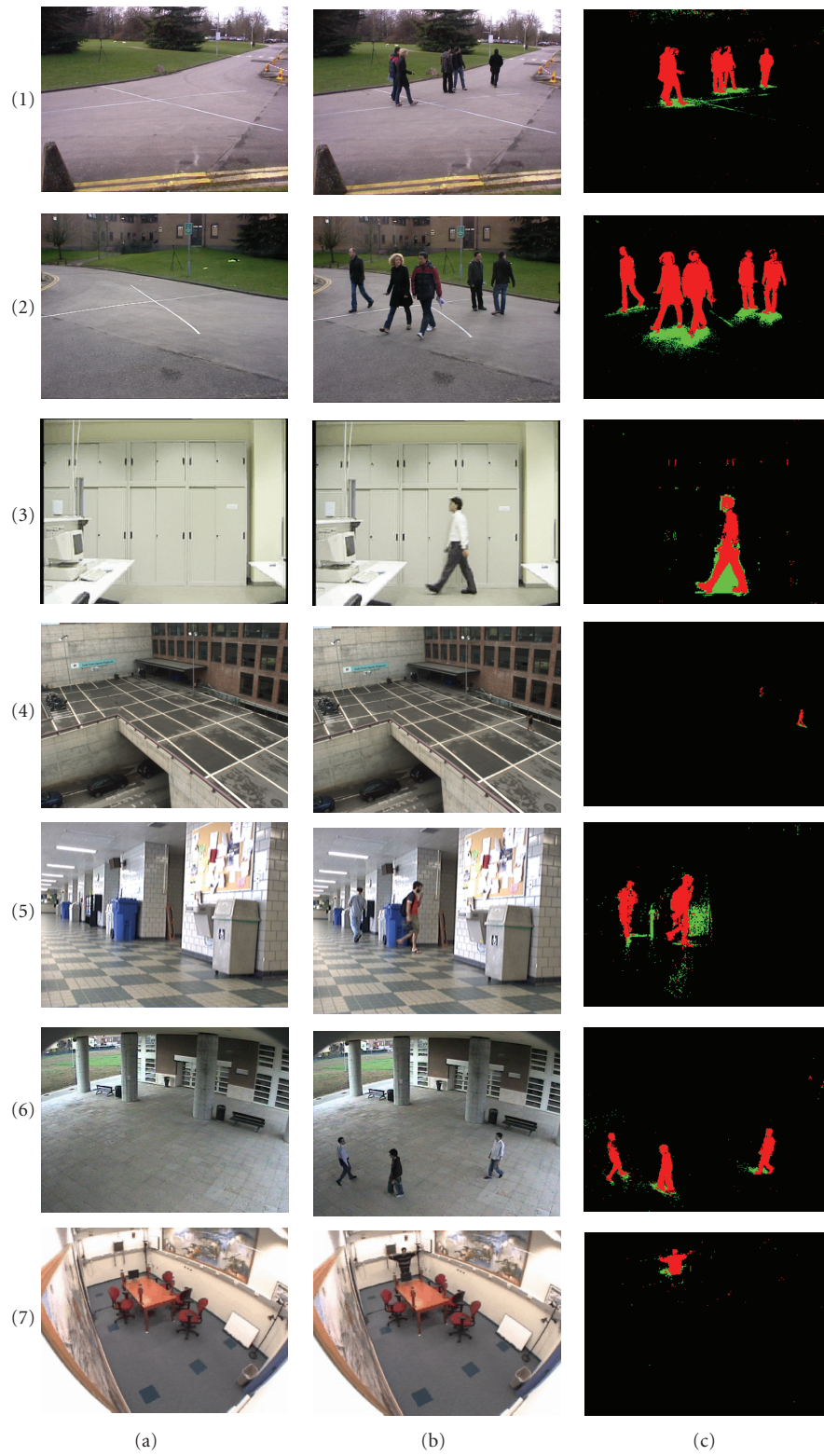


FIGURE 4: Sample visual results of our background subtraction algorithm in various environment. (a) Background Image, (b) Current Image, and (c) Foreground (red) /Shadows (green) /Background (black) detection. (1) PETS 2009 View 7, (2) PETS 2009 View 8, (3) ATON (Laboratory), (4) ISELAB (ETSE Outdoor), (5) LVS (HallwayI), (6) VSSN, and (7) ATON (Intelligentroom).

3.2. Classification Process. Our classification rules consist of two steps.

Step One. Pixels that have strong dissimilarity with the background are classified directly as foreground, in the case when the following rule expression is equal to 1 (TRUE):

$$\begin{aligned} \text{Fr}(\mathbf{x}) = & T\theta\left(\Delta\theta\left(\mathbf{p}^{bg}(\mathbf{x}), \mathbf{p}^f(\mathbf{x})\right), \gamma^\theta\right) \\ & \cap TI\left(\Delta I\left(\mathbf{p}^{bg}(\mathbf{x}), \mathbf{p}^f(\mathbf{x})\right), \gamma^I\right), \end{aligned} \quad (9)$$

where γ^θ and γ^I are experimental scale factors. Otherwise, when (9) is not TRUE, the classification has to be done in the following step.

Step Two. This step consists of two test rules. One verifies a test pixel for the shadow class (10) and another verifies for the foreground class (11):

$$\begin{aligned} \text{Sh}(\mathbf{x}) = & T\mu I\left(\mu I\left(\left(\mathbf{p}^{bg}(\mathbf{x} + \mathbf{w}), \mathbf{p}^f(\mathbf{x} + \mathbf{w})\right), \gamma_I T^I(\mathbf{x})\right), k_F^I\right) \\ & \cap \left(\left|\mathbf{p}^{bg}(\mathbf{x})\right| > \left|\mathbf{p}^f(\mathbf{x})\right|\right) \\ & \cap \left(1 - T\eta\theta\left(\eta\theta\left(\left(\mathbf{p}^{bg}(\mathbf{x} + \mathbf{w}), \mathbf{p}^f(\mathbf{x} + \mathbf{w})\right), \gamma_\theta T^\theta(\mathbf{x})\right), k_S^\theta\right)\right) \\ & \cap \left(1 - T\mu I\left(\mu I\left(\left(\mathbf{p}^{bg}(\mathbf{x} + \mathbf{w}), \mathbf{p}^f(\mathbf{x} + \mathbf{w})\right), \gamma_S T^I(\mathbf{x})\right), k_S^I\right)\right), \end{aligned} \quad (10)$$

$$\begin{aligned} \text{Fr}(\mathbf{x}) = & T\mu I\left(\mu I\left(\left(\mathbf{p}^{bg}(\mathbf{x} + \mathbf{w}), \mathbf{p}^f(\mathbf{x} + \mathbf{w})\right), \gamma_I T^I(\mathbf{x})\right), k_F^I\right) \\ & \cap (1 - \text{Sh}(\mathbf{x})). \end{aligned} \quad (11)$$

The rest of the pixels that are not classified as shadow or foreground pixels must be classified as background pixels. Figure 1 illustrates the classification regions. All the implemented thresholds were obtained on the base of a tuning process with different video sequences ($\gamma^\theta = 10^\circ$, $\gamma^I = 55$, $\gamma_I = 10$, $\gamma_\theta = 2^\circ$, $\gamma_S = 80$ and $K_F^I = K_S^\theta = K_S^I = 1$).

3.3. Model Updating. In order to maintain the stability of the background model through the time, the model needs to be dynamically updated. As it was explained before, only the RCs have to be updated. The update process is done at every frame, but only in the case when the updated pixels are classified as a background. The model is updated as follows:

$$p_c^{bg}(\mathbf{x}, t) = \beta p_c^{bg}(\mathbf{x}, t-1) + (1-\beta) p_c^f(\mathbf{x}, t), \quad c \in \{R, G, B\}, \quad (12)$$

where $(0 < \beta < 1)$ is the updated rate. Due to our experiments the value of this parameter has to be $\beta = 0.45$.

4. Experimental Results

In this section we present the performance of our approach in terms of quantitative and qualitative results applied to 5

well-known datasets taken from 7 different video sequences: PETS 2009 (<http://www.cvg.rdg.ac.uk/> (View 7 and 8)), ATON (<http://cvrr.ucsd.edu/aton/shadow/> (Laboratory and Intelligentroom)), ISELAB (<http://iselab.cvc.uab.es> (ETSE Outdoor)), LVSN (<http://vision.gel.ulaval.ca/CastShadows/> (HallwayI)), and VSSN, (http://mmc36.informatik.uni-augsburg.de/VSSN06_OSAC/).

Quantitative Results. We have applied our proposed algorithm in several indoor and outdoor video scenes. Ground-truth masks have been manually extracted to numerically evaluate and compare the performance of our proposed technique with respect to most similar state-of-the-art approaches [6–9]. Two metrics were considered to evaluate the segmentation results, namely, False Positive Error (FPE) and False Negative Error (FNE). FPE means that the background pixels were set as Foreground while FNE indicates that foreground pixels were identified as Background. We show this comparison in terms of accuracy in Figure 2:

$$\text{Error}(\%) = \frac{\text{No. of misclassification pixels}}{\text{No. of correct foreground pixels}} \times 100\%. \quad (13)$$

Qualitative Results. Figure 3 shows a visual comparison between our techniques and some well-known methods. It can be seen that our method performs better in terms of camouflage areas segmentation and suppressing strong shadows. In Figure 4 also visual results are shown. In this case we have applied our method in several sequences. It can be seen that the foreground objects are detected without shadows, in such a way preserving their shape properly.

5. Conclusions

This paper proposes an efficient background subtraction technique which overcomes difficulties like illumination changes and moving shadows. The main novelty of our method is the incorporation of two discriminative similarity measures based on angular and Euclidean distance patterns in local neighborhoods. Such patterns are used to improve foreground detection in the presence of moving shadows and strong similarities in color between background and foreground. Experimental results over a collection of public and own datasets of real image sequences demonstrate the effectiveness of the proposed technique. The method shows an excellent performance in comparison with other methods. Most recent approaches are based on very complex models designed to achieve an extremely effective classification; however these approaches become unfeasible for real-time applications. Alternatively, our proposed method exhibits low computational and space complexities that make our proposal very appropriate for real-time processing in surveillance systems with low-resolution cameras or Internet webcams.

Acknowledgments

This work has been supported by the Spanish Research Programs Consolider-Ingenio 2010:MIPRCV (CSD200700018) and Avanza I+D ViCoMo (TSI-020400-2009-133) and by the Spanish projects TIN2009-14501-C02-01 and TIN2009-14501-C02-02.

References

- [1] M. Karaman, L. Goldmann, D. Yu, and T. Sikora, "Comparison of static background segmentation methods," in *Visual Communications and Image Processing*, vol. 5960 of *Proceedings of SPIE*, no. 4, pp. 2140–2151, 2005.
- [2] M. Piccardi, "Background subtraction techniques: a review," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC '04)*, vol. 4, pp. 3099–3104, The Hague, The Netherlands, October 2004.
- [3] A. McIvor, "Background subtraction techniques," in *Proceedings of the International Conference on Image and Vision Computing*, Auckland, New Zealand, 2000.
- [4] A. Prati, I. Mikic, M. M. Trivedi, and R. Cucchiara, "Detecting moving shadows: algorithms and evaluation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 7, pp. 918–923, 2003.
- [5] G. Obinata and A. Dutta, *Vision Systems: Segmentation and Pattern Recognition*, I-TECH Education and Publishing, Vienna, Austria, 2007.
- [6] I. Haritaoglu, D. Harwood, and L. S. Davis, "W4: real-time surveillance of people and their activities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 809–830, 2000.
- [7] T. Hoprasert, D. Harwood, and L. S. Davis, "A statistical approach for real-time robust background subtraction and shadow detection," in *Proceedings of the 7th IEEE International Conference on Computer Vision, Frame Rate Workshop (ICCV '99)*, vol. 4, pp. 1–9, Kerkyra, Greece, September 1999.
- [8] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Real-time foreground-background segmentation using codebook model," *Real-Time Imaging*, vol. 11, no. 3, pp. 172–185, 2005.
- [9] C. Stauffer and W. E. L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 747–757, 2000.
- [10] S. J. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler, "Tracking groups of people," *Computer Vision and Image Understanding*, vol. 80, no. 1, pp. 42–56, 2000.
- [11] R. Cucchiara, C. Grana, M. Piccardi, A. Prati, and S. Sirotti, "Improving shadow suppression in moving object detection with HSV color information," in *Proceedings of the IEEE Intelligent Transportation Systems Proceedings*, pp. 334–339, Oakland, Calif, USA, August 2001.
- [12] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: principles and practice of background maintenance," in *Proceedings of the 7th IEEE International Conference on Computer Vision (ICCV '99)*, vol. 1, pp. 255–261, Kerkyra, Greece, September 1999.
- [13] A. Elgammal, D. Harwood, and L. S. Davis, "Nonparametric background model for background subtraction," in *Proceedings of the European Conference on Computer Vision (ECCV '00)*, pp. 751–767, Dublin, Ireland, 2000.
- [14] A. Mittal and N. Paragios, "Motion-based background subtraction using adaptive kernel density estimation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, vol. 2, pp. 302–309, Washington, DC, USA, July 2004.
- [15] Y.-T. Chen, C.-S. Chen, C.-R. Huang, and Y.-P. Hung, "Efficient hierarchical method for background subtraction," *Pattern Recognition*, vol. 40, no. 10, pp. 2706–2715, 2007.
- [16] L. Li, W. Huang, I. Y.-H. Gu, and Q. Tian, "Statistical modeling of complex backgrounds for foreground object detection," *IEEE Transactions on Image Processing*, vol. 13, no. 11, pp. 1459–1472, 2004.
- [17] J. Zhong and S. Sclaroff, "Segmenting foreground objects from a dynamic textured background via a robust Kalman filter," in *Proceedings of the 9th IEEE International Conference on Computer Vision (ICCV '03)*, pp. 44–50, Nice, France, October 2003.