

# VERSATILE BAYESIAN CLASSIFIER FOR MOVING OBJECT DETECTION BY NON-PARAMETRIC BACKGROUND-FOREGROUND MODELING

*Carlos Cuevas, Raúl Mohedano, and Narciso García*

## ABSTRACT

Along the recent years, several moving object detection strategies by non-parametric background-foreground modeling have been proposed. To combine both models and to obtain the probability of a pixel to belong to the foreground, these strategies make use of Bayesian classifiers. However, these classifiers do not allow to take advantage of additional prior information at different pixels. So, we propose a novel and efficient alternative Bayesian classifier that is suitable for this kind of strategies and that allows the use of whatever prior information. Additionally, we present an effective method to dynamically estimate prior probability from the result of a particle filter-based tracking strategy.

**Index Terms**— Moving object detection, Bayesian classifier, prior probability estimation, background-foreground modeling.

## 1. INTRODUCTION

The recent huge proliferation of electronic devices with camera platforms has resulted in an important demand for new and efficient computer vision applications [1]. In these applications, the moving object detection is a key step for high level analysis tasks such as segmentation, event analysis, or tracking.

On the one hand, simplest moving object detection strategies try to be fast and to reduce memory requirements. However, they do not provide satisfactory results in complex scenarios with dynamic backgrounds (containing rain, snow, waving flags or trees, etc.) and depend on several thresholds than should be manually set according to the characteristics of the analyzed sequence [2]. On the other hand, several multimodal alternatives have been also proposed, which are able to improve the quality of the detections in scenarios with non-static backgrounds by modeling multiple states for each pixel [3].

Among multimodal strategies, non-parametric based methods have probably been those that have drawn the most attention of the researchers, as they are able to provide very high quality detections even in environments where the pixel variations can not be described with other multimodal methods. These strategies do not consider the values of the pixels as a particular distribution and build a probabilistic representation of the observations using a recent sample of values for each pixel [4].

To improve the quality of the detections in scenarios where moving objects and foreground have similar characteristics, some non-parametric-based proposals estimate not only a background

density function but also a foreground model [5]. These strategies usually make use of spatio-temporal reference data to avoid false detections resulting from small displacements of the background (for example, in sequences recorded with non-stabilized cameras) and to facilitate the foreground modeling [1]. Finally, to estimate the probability of a pixel to belong to the foreground they use a Bayesian classifier where the background model, the foreground model, and the foreground and background prior probabilities are combined [6]. However, typical Bayesian classifiers do not allow to include prior information at different spatial positions. Therefore, these proposals can not take advantage of additional prior information resulting from the analysis of each image.

In this paper we propose a novel and efficient alternative Bayesian classifier which, unlike those reported before, allows the use of prior information obtained from whatever source of information and depending on the spatial position of each pixel. Moreover, we present an effective method that dynamically estimates the foreground prior probability from the information provided by a particle filter-based tracking strategy. As a result of the combination of background and foreground models with the estimated prior probabilities, the quality of the detections is improved, as more accurate and compact moving regions are obtained.

## 2. SPATIO-TEMPORAL NON-PARAMETRIC MODELING

Let us consider a pixel  $p^n$  in the current image  $I^n$ , at time  $n$ , defined as a  $(D + 2)$ -dimensional vector,  $\mathbf{x}^n = ((\mathbf{c}^n)^T, (\mathbf{s}^n)^T)^T \in \mathbb{R}^{D+2}$ , where  $\mathbf{c}^n \in \mathbb{R}^D$  is a vector containing appearance characteristics of  $p^n$  (e.g. color, gradient, depth, etc.) and  $\mathbf{s}^n = (h^n, w^n) \in \mathbb{R}^2$  is a vector containing its spatial coordinates (rows and columns). Using  $(D + 2)$ -dimensional spatio-temporal samples, extracted from previous images into a spatial neighborhood around the spatial position of  $p^n$ , the probability density function of background,  $\beta$ , and foreground  $\phi$ , can be very successfully non-parametrically estimated [5].

Once both background and foreground have been modeled, the probability of  $p^n$  to belong to the foreground class can be efficiently estimated [6] using Bayes' theorem:

$$Pr(\phi|\mathbf{x}^n) = \frac{Pr(\phi)p(\mathbf{x}^n|\phi)}{Pr(\phi)p(\mathbf{x}^n|\phi) + Pr(\beta)p(\mathbf{x}^n|\beta)}, \quad (1)$$

where  $Pr(\phi)$  is the foreground prior probability,  $Pr(\beta) = 1 - Pr(\phi)$  is the background prior probability, and  $p(\mathbf{x}^n|\beta)$  and  $p(\mathbf{x}^n|\phi)$  are, respectively, the estimated background and foreground density functions.

---

This work has been partially supported by the Ministerio de Ciencia e Innovación of the Spanish Government under project TEC2010-20412 (Enhanced 3DTV).

## 2.1. Background modeling

Let us consider a set of  $N_\beta$  background reference samples,  $\mathbf{x}_\beta^i = (\mathbf{c}_\beta^i, \mathbf{s}_\beta^i)$ , obtained from  $T_\beta$  previous images ( $T_\beta \leq N_\beta$ ) into a spatial neighborhood around the coordinates of  $p^n$ . Using Gaussian kernels, the probability density function that  $\mathbf{x}^n$  belongs to the image background,  $\beta$ , can be non-parametrically estimated [5] as

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

where  $\Sigma_\beta$  is a symmetric definite  $(D+2) \times (D+2)$  bandwidth matrix. Looking for a trade-off between computational efficiency and quality [7],  $\Sigma_\beta$  is defined as

$$\Sigma_\beta = \text{diag}(\sigma_{\beta_1}^2, \sigma_{\beta_2}^2 \dots \sigma_{\beta_D}^2, \sigma_{\beta_H}^2, \sigma_{\beta_W}^2), \quad (3)$$

where the first  $D$  components determine the bandwidth of the appearance components and the two last determine the spatial bandwidth of the Gaussian kernels.

## 2.2. Foreground modeling

At first, the probability density function that  $p^n$  belongs to the foreground is uniform. However, this probability increases if moving objects have been previously detected around  $p^n$ . Consequently, an adequate foreground model can be estimated as a mixture of a uniform function,  $\gamma$ , and a Gaussian kernel density estimation [6],

$$p(\mathbf{x}^n|\phi) = \alpha\gamma + \frac{(1-\alpha)}{N_\phi(2\pi)^{\frac{D+2}{2}}} \times \sum_{i=1}^{N_\phi} \prod_{j=1}^{D+2} \frac{1}{(\Sigma_\phi(j, j))^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \frac{(\mathbf{x}^n(j) - \mathbf{x}_\phi^i(j))^2}{\Sigma_\phi(j, j)}\right), \quad (4)$$

where  $\mathbf{x}_\phi^i = (\mathbf{c}_\phi^i, \mathbf{s}_\phi^i)$  are the  $N_\phi$  samples classified as foreground along the previous  $T_\phi$  images,  $\alpha$  is a mixture factor, and  $\Sigma_\phi$  is the bandwidth matrix for the Gaussian kernels which, similarly to the bandwidth matrices used in the background modeling and for similar reasons, are defined as

$$\Sigma_\phi = \text{diag}(\sigma_{\phi_1}^2, \sigma_{\phi_2}^2 \dots \sigma_{\phi_D}^2, \sigma_{\phi_H}^2, \sigma_{\phi_W}^2) \quad (5)$$

where the first  $D$  components determine the bandwidth of the appearance components and the two last determine the spatial bandwidth of the Gaussian kernels.

Additionally, using an efficient particle filter-based tracking strategy [8] we update the spatial positions of previously detected moving objects. This update allows to select much more appropriate foreground bandwidth matrices (much more fitted to the size of moving objects), improving the quality of the foreground modeling and reducing its computational cost. Moreover, the amount of predicted particles provided by this filter is higher at the positions of the image where the moving regions are most probable to be allocated in the future (according the trajectory and the speed of these regions). So, the distribution of predicted particles on the images can be used as prior information.

## 3. ALTERNATIVE BAYESIAN CLASSIFIER

Initially, if the only available information is that provided by the estimated background and foreground models, the prior probabilities of both classes should be equal,  $Pr(\beta) = Pr(\phi) = \frac{1}{2}$ . However, additional information is normally available (i.e. the spatial positions of previously detected moving objects, their trajectories, or their speeds) and can provide clues about areas of the images where moving objects will most probably appear. As the used foreground modeling is estimated from spatio-temporal information of previously detected foreground pixels, it takes into account prior information concerning the location of these moving objects in the scene. Nevertheless, using a classifier as described in (1), where the prior probability is a constant value at any spatial position, is not possible to include additional prior information at each pixel (as, for example, information relative to the trajectory and the speed of previously detected foreground regions).

To make use of this information or any other available data, in this paper we propose an innovative and alternative classifier that allows the use of prior probability depending on the spatial position of each pixel and that, moreover, can be estimated from whatever source of information.

The proposed classifier obtains the probability of  $p^n$  to belong to the foreground class as:

$$\begin{aligned} Pr(\phi|\mathbf{x}^n) &= Pr(\phi|\mathbf{c}^n, \mathbf{s}^n) = \frac{p(\mathbf{c}^n, \phi|\mathbf{s}^n)}{p(\mathbf{c}^n|\mathbf{s}^n)} = \\ &= \frac{Pr(\phi|\mathbf{s}^n)p(\mathbf{c}^n|\mathbf{s}^n, \phi)}{p(\mathbf{c}^n, \phi|\mathbf{s}^n) + p(\mathbf{c}^n, \beta|\mathbf{s}^n)} = \\ &= \frac{Pr(\phi|\mathbf{s}^n)p(\mathbf{c}^n|\mathbf{s}^n, \phi)}{Pr(\phi|\mathbf{s}^n)p(\mathbf{c}^n|\mathbf{s}^n, \phi) + Pr(\beta|\mathbf{s}^n)p(\mathbf{c}^n|\mathbf{s}^n, \beta)}. \end{aligned} \quad (6)$$

In this equation,  $Pr(\phi|\mathbf{s}^n)$  and  $Pr(\beta|\mathbf{s}^n) = 1 - Pr(\phi|\mathbf{s}^n)$  are the prior probabilities of foreground and background (both space-dependent), and  $p(\mathbf{c}^n|\mathbf{s}^n, \phi)$  and  $p(\mathbf{c}^n|\mathbf{s}^n, \beta)$  are the foreground and background estimated models, conditioned in space. These conditioned probabilities can be expressed as

$$p(\mathbf{c}^n|\mathbf{s}^n, \phi) = \frac{p(\mathbf{x}^n|\phi)}{p(\mathbf{s}^n|\phi)}, \quad (7)$$

$$p(\mathbf{c}^n|\mathbf{s}^n, \beta) = \frac{p(\mathbf{x}^n|\beta)}{p(\mathbf{s}^n|\beta)}, \quad (8)$$

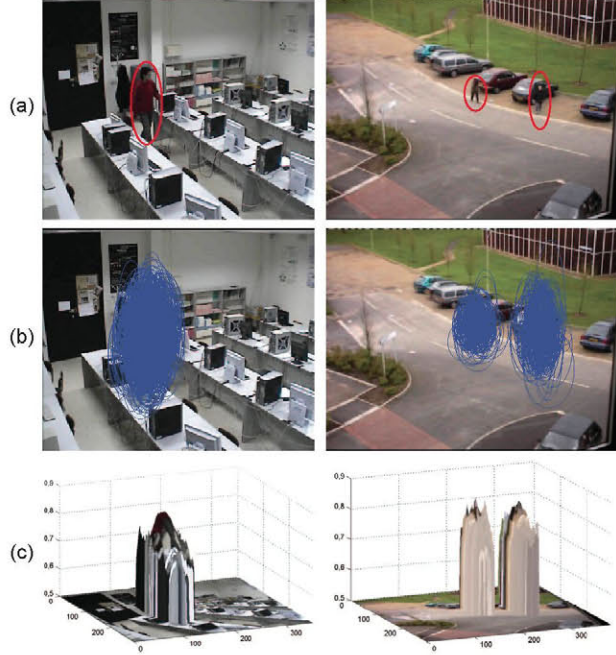
where  $p(\mathbf{s}^n|\phi)$  is the marginalisation of the estimated foreground likelihood,  $p(\mathbf{x}^n|\phi)$ , over the appearance components,

$$\begin{aligned} p(\mathbf{s}^n|\phi) &= \alpha\gamma' + \\ &+ \frac{(1-\alpha)}{N_\phi 2\pi} \sum_{i=1}^{N_\phi} \frac{1}{|\Sigma_{\phi, \mathbf{s}}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (\mathbf{s}^n - \mathbf{s}_\phi^i)^T \Sigma_{\phi, \mathbf{s}}^{-1} (\mathbf{s}^n - \mathbf{s}_\phi^i)\right) \end{aligned} \quad (9)$$

and  $p(\mathbf{s}^n|\beta)$  is the marginalisation of the estimated background likelihood,  $p(\mathbf{x}^n|\beta)$ , over the appearance components,

$$\begin{aligned} p(\mathbf{s}^n|\beta) &= \\ &= \frac{1}{N_\beta 2\pi} \sum_{i=1}^{N_\beta} \frac{1}{|\Sigma_{\beta, \mathbf{s}}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (\mathbf{s}^n - \mathbf{s}_\beta^i)^T \Sigma_{\beta, \mathbf{s}}^{-1} (\mathbf{s}^n - \mathbf{s}_\beta^i)\right). \end{aligned} \quad (10)$$

In these equations,  $\Sigma_{\phi, \mathbf{s}}$  and  $\Sigma_{\beta, \mathbf{s}}$  are the  $2 \times 2$  dimensional matrices that determine the spatial bandwidth of the kernels,  $\alpha \in [0, 1]$



**Fig. 1.** (a) Point estimation of the moving objects in the original images. (b) Propagated particles. (c) Three dimensional representation of  $Pr_\phi(\mathbf{s}^n)$ .

is the same mixture factor used in (4), and  $\gamma'$  is a constant density of a uniform random variable in the spatial set of components.

Since the bandwidth of the Gaussian kernels is determined by diagonal matrices,  $\Sigma_{\phi,s}$  and  $\Sigma_{\beta,s}$  can be derived directly from  $\Sigma_\phi$  and  $\Sigma_\beta$ . Therefore, equations (7) and (8) are obtained without additional computational effort, as they can be computed simultaneously to  $p(\mathbf{x}^n|\phi)$  and  $p(\mathbf{x}^n|\beta)$ .

#### 4. ESTIMATION OF PRIOR PROBABILITIES

The proposed Bayesian classifier allows the combination of the estimated background and foreground models with pixel-wise prior information such as information on positions, trajectories and velocities of previously detected moving objects. On the one hand, we propose to use the information corresponding to the position of previously detected moving regions that is obtained from the marginal probabilities defined in (9) and (10). On the other hand, we also use the trajectories and the speed of these regions that are provided by the predictions resulting from the application of the particle filter. The combination of both types of information improves the detection quality of the classifier.

In the Bayesian classifier described in (1) the background and foreground models are directly applied. Hence, the position of previously detected moving objects is taken into account. However, the alternative classifier described in (6) uses the conditional probabilities of foreground and background and, therefore, this information is not considered. Therefore, to take into account this information, the prior probabilities of this equation should contemplate the marginal probabilities described in (9) and (10).

On the other hand, as a result of the particle filter used to update the position of previous detections a set of predicted particles is obtained. The amount of these particles will be higher in regions

of the image where, depending on the trajectory and the speed of the previously detected moving objects, these objects will most probably appear. Therefore, the information provided by these particles can also be used as prior information. Figure 1 shows some results provided by this filter (the point estimations in Fig.1.a and the propagated particles in Fig.1.b) in two different scenarios.

To make use of both the marginal probabilities and the predictions provided by the particle filter we propose the use of prior probabilities defined as

$$Pr(\phi|\mathbf{s}^n) = \frac{Pr_\phi(\mathbf{s}^n)p(\mathbf{s}^n|\phi)}{Pr_\phi(\mathbf{s}^n)p(\mathbf{s}^n|\phi) + Pr_\beta(\mathbf{s}^n)p(\mathbf{s}^n|\beta)}, \quad (11)$$

$$Pr(\beta|\mathbf{s}^n) = \frac{Pr_\beta(\mathbf{s}^n)p(\mathbf{s}^n|\beta)}{Pr_\phi(\mathbf{s}^n)p(\mathbf{s}^n|\phi) + Pr_\beta(\mathbf{s}^n)p(\mathbf{s}^n|\beta)}, \quad (12)$$

where  $Pr_\phi(\mathbf{s}^n)$  is the probability obtained from the predictions of the particle filter and  $Pr_\beta(\mathbf{s}^n) = 1 - Pr_\phi(\mathbf{s}^n)$  is the complementary probability. Thus, the alternative classifier defined in (6) can be formulated as

$$Pr(\phi|\mathbf{x}^n) = \frac{Pr_\phi(\mathbf{s}^n)p(\mathbf{x}^n|\phi)}{Pr_\phi(\mathbf{s}^n)p(\mathbf{x}^n|\phi) + Pr_\beta(\mathbf{s}^n)p(\mathbf{x}^n|\beta)}. \quad (13)$$

Therefore, to use the proposed classifier, in addition to the estimated background and foreground models, only  $Pr_\phi(\mathbf{s}^n)$  should be computed.

In image areas that are not covered by predicted particles there is not a priori information on the speed or the trajectory of moving objects. Consequently, in these areas the prior probabilities must be established as  $Pr_\phi(\mathbf{s}^n) = Pr_\beta(\mathbf{s}^n) = \frac{1}{2}$ . However, in areas covered by predicted particles the value of  $Pr_\phi(\mathbf{s}^n)$  must be higher than  $\frac{1}{2}$  (with lower or higher value depending on the amount of particles covering each pixel and the distance between the centers of the particles and these pixels). Based on these criteria  $Pr_\phi(\mathbf{s}^n)$  has been defined as

$$Pr_\phi(\mathbf{s}^n) = \begin{cases} \frac{1}{2}, & N_p \leq N_n \\ \frac{1}{N_p} \sum_{i=1}^{N_p} G_i(\mathbf{s}^n), & N_p > N_n \end{cases} \quad (14)$$

where  $N_p$  is the number of predicted particles on the analyzed coordinates,  $N_n$  is a threshold to avoid the influence of noisy predictions, and  $G_i(\mathbf{s}^n)$  is the profile of the  $i$ -th particle on that position. The predicted particles provided by the filter are defined as ellipses and we have decided to set their profiles as normalized bivariate Gaussians defined as

$$G_i(\mathbf{s}^n) = \exp\left(-\frac{1}{2} \left( \frac{(\mathbf{s}^n(1) - h_i^n)^2}{(\eta a_i^n)^2} + \frac{(\mathbf{s}^n(2) - w_i^n)^2}{(\eta b_i^n)^2} \right)\right), \quad (15)$$

where  $(h_i^n, w_i^n)$  are the coordinates of the center of each ellipse,  $(a_i^n, b_i^n)$  are their axes, and  $\eta$  is a factor that sets the values of the Gaussians at the contour of the ellipses as  $\frac{1}{2}$ . Therefore, its value is  $\eta = (2 \ln(2))^{-\frac{1}{2}} \approx 0.85$ .

The last row of images in Fig.1 shows two three dimensional representations of  $Pr_\phi(\mathbf{s}^n)$ , obtained from the sets of predicted particles represented in the second row of the figure.

## 5. RESULTS

The proposed strategy has been tested in several indoor and outdoor scenarios recorded with non-stabilized cameras and containing critical situations such as complex and dynamic backgrounds, shadows, and multiple moving objects similar to background regions. These sequences have been extracted from the PETS database [9], the Wallflower database [10], and our own database [11].

We have used a buffer of  $T_\beta = 150$  images and a buffer of  $T_\phi = 10$  images to model the background and the foreground, respectively. The appearance information of the pixels are their RGB color components, so  $D = 3$ . Moreover, we have dynamically estimated adequate values of  $\Sigma_\beta$  and  $\Sigma_\phi$ .

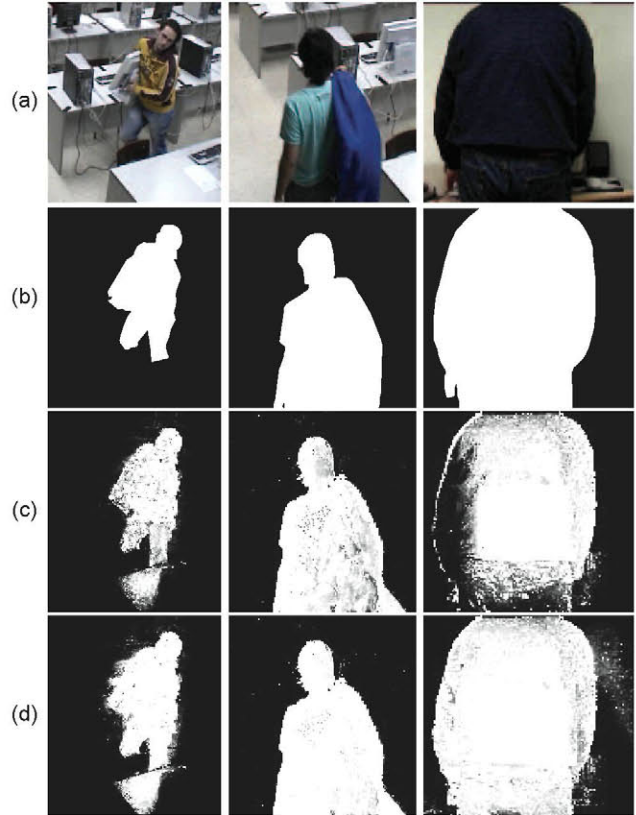
The obtained results have demonstrate that the use of the estimated prior probabilities results in higher quality results, with more compact and better defined moving objects, mainly in sequences where the moving objects remain static temporally. Figure 2 shows some of the results obtained for three scenarios with different characteristics. The first row of images (Fig.2.a) presents the original images and the second row (Fig.2.b) shows the corresponding ground truth. The detections obtained by using equal foreground and background prior probabilities are depicted in the second row of images (Fig.2.c), while the detections obtained with the estimated prior information is shown in the last row (Fig.2.d). These results allow us to appreciate that the application of the proposed strategy decreases the number of misdetections, consequently, the obtained detections are more compact.

## 6. CONCLUSIONS

We have presented a novel and efficient Bayesian classifier that is suitable for moving object detection strategies by spatio-temporal background-foreground non-parametric modeling. This classifier, unlike those used in previous works, allows to make use of whatever prior information adapted to each pixel location. Additionally, we have proposed a strategy to dynamically obtain prior probabilities from a particle filter-based tracking strategy. Moreover, we have demonstrated that applying this probability to the proposed Bayesian classifier the quality of the detections improves very significantly.

## 7. REFERENCES

- [1] Y. Sheikh, O. Javed, and T. Kanade, "Background subtraction for freely moving cameras," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 1219–1225.
- [2] M. Piccardi, "Background subtraction techniques: a review," in *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, 2005, vol. 4, pp. 3099–3104.
- [3] M. Cristani, M. Farenzena, D. Bloisi, and V. Murino, "Background subtraction for automated multisensor surveillance: a comprehensive review," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, pp. 43, 2010.
- [4] T. Tanaka, A. Shimada, R. Taniguchi, T. Yamashita, and D. Arita, "Towards robust object detection: integrated background modeling based on spatio-temporal features," *Computer Vision-ACCV 2009*, pp. 201–212, 2010.
- [5] Y. Sheikh and M. Shah, "Bayesian modeling of dynamic scenes for object detection," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 11, pp. 1778–1792, 2005.
- [6] C. Cuevas, R. Mohedano, F. Jaureguizar, and N. García, "High-quality real-time moving object detection by non-parametric segmentation," *Electronics Letters*, vol. 46, no. 13, pp. 910–911, 2010.
- [7] N. Martel-Brisson and A. Zaccarin, "Unsupervised approach for building non-parametric background and foreground models of scenes with significant foreground activity," in *Proceeding of the 1st ACM workshop on Vision networks for behavior analysis*. ACM, 2008, pp. 93–100.
- [8] M. Nieto, C. Cuevas, and L. Salgado, "Measurement-based reclustering for multiple object tracking with particle filters," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*. IEEE, 2010, pp. 4097–4100.
- [9] Computational Vision Group. University of Reading, "Pets: Performance evaluation of tracking and surveillance," <http://www.cvg.rdg.ac.uk/>.
- [10] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: principles and practice of background maintenance," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, 1999, vol. 1, pp. 255–261 vol.1. <http://research.microsoft.com/en-us/um/people/jckrumm/wallflower/testimages.html>.
- [11] Grupo de Tratamiento de Imágenes. Universidad Potitécnica de Madrid, "Lab database," <http://www.gti.ssr.upm.es/data/>.



**Fig. 2.** (a) Original images. (b) Ground truth. (c) Final detections with equal background and foreground prior probabilities. (d) Final detections with the estimated prior probabilities.