Adaptable Bayesian classifier for spatiotemporal nonparametric moving object detection strategies

Carlos Cuevas,* Raúl Mohedano, and Narciso García

E.T.S.I. Telecomunicación, Universidad Politécnica de Madrid (UPM), 28040, Madrid, Spain
**Corresponding author: ccr@gti.ssr.upm.es*

Received April 26, 2012; revised June 25, 2012; accepted June 25, 2012;
posted June 26, 2012 (Doc. ID 167504); published July 25, 2012

Electronic devices endowed with camera platforms require new and powerful machine vision applications, which commonly include moving object detection strategies. To obtain high-quality results, the most recent strategies estimate nonparametrically background and foreground models and combine them by means of a Bayesian classifier. However, typical classifiers are limited by the use of constant prior values and they do not allow the inclusion of additional spatiodependent prior information. In this Letter, we propose an alternative Bayesian classifier that, unlike those reported before, allows the use of additional prior information obtained from any source and depending on the spatial position of each pixel.

OCIS codes: 100.2960, 150.0150.

Recently, the number of electronic devices endowed with camera platforms has increased very significantly [1]. Consequently, the demand for image processing applications required by these devices has also increased [2]. To perform high level analysis tasks, these applications include as a first step moving object detection strategies [3]. The simplest detection strategies aim to be fast and to reduce memory requirements [4]. However, they do not provide satisfactory results in complex scenarios and depend on several thresholds [5]. To improve the quality of the detections and avoid the use of thresholds, several multimodal strategies have been also proposed [6]. These strategies are able to model multiple alternating states for each pixel, and prove themselves effective in most practical scenarios.

Among multimodal strategies, nonparametric methods have drawn the most attention from the researchers [7]. These strategies do not consider the values of the pixels as a particular distribution, and build instead probabilistic representations of the underlying process from sets of recent samples for each pixel [8]. In this way, they are able to provide very high quality detections in environments where other methods cannot correctly describe pixel variations.

To avoid false detections resulting from small displacements of the background, these strategies use spatiotemporal reference data in the modeling [9]. Furthermore, to improve the quality of the detections in scenarios where moving objects and background have similar characteristics [10], most recent nonparametric proposals estimate not only a background density function but also a foreground model, combining both models by means of a Bayesian classifier [11]. However, typically used Bayesian classifiers only use prior probabilities defined constantly over the whole image extent [12]. Consequently, they do not allow the inclusion of additional information sources expressing the prior confidence in each pixel location to be occupied by a moving object, which are commonly available as the result of auxiliary processes of many detection strategies [13].

To overcome this important limitation, we propose a novel and efficient adaptable Bayesian classifier. This classifier, unlike those reported before in the literature, allows the natural integration of additional prior information obtained from any source and depending on the spatial position of each pixel. In this way, it improves the quality of the detections when the estimated background and foreground models do not discriminate adequately the moving objects from the background regions.

The proposed adaptable Bayesian classifier was originally designed for moving object detection strategies, hence the notation deployed throughout this Letter. However, many other computer vision and optics applications using Bayesian approaches could also naturally benefit from the flexibility and spatial information fusion capabilities of our proposal (adapted correspondingly to the application under study). As an example of those potentially benefited applications we could cite: anomaly detection in infrared [14] and hyperspectral [15] imaging, detection and characterization of discrete objects in astrophysics and cosmology [16], or salient object detection in static images [17].

Let p^n be a pixel in the current image I^n, at time n, defined as a $(D+2)$-dimensional vector $\mathbf{x}^n = ((\mathbf{a}^n)^T, (\mathbf{s}^n)^T)^T \in \mathbb{R}^{D+2}$, where $\mathbf{a}^n \in \mathbb{R}^D$ is a vector containing appearance information of p^n (e.g., color, gradient, depth) and $\mathbf{s}^n = (r^n, c^n) \in \mathbb{R}^2$ is a vector containing its spatial coordinates (rows and columns).

Using $(D+2)$-dimensional spatiotemporal reference samples and applying Gaussian kernels, nonparametric moving object detection strategies [9] commonly construct the nonparametric models for the background, β, and for the foreground, ϕ, as follows:

$$M_\beta(\mathbf{x}^n) = M_\beta(\mathbf{a}^n, \mathbf{s}^n) = \frac{1}{N_\beta} \sum_{i=1}^{N_\beta} \frac{|\Sigma_\beta|^{-1/2}}{(2\pi)^{\frac{D+2}{2}}} \times \exp\left(-\frac{1}{2}(\mathbf{x}^n - \mathbf{x}^i_\beta)^T \left(\Sigma_\beta\right)^{-1} (\mathbf{x}^n - \mathbf{x}^i_\beta)\right), \quad (1)$$

0146-9592/12/153159-03$15.00/0

$$M_\phi(\mathbf{x}^n) = M_\phi(\mathbf{a}^n, \mathbf{s}^n) = \alpha\gamma + \frac{(1-\alpha)}{N_\phi} \sum_{i=1}^{N_\phi} \frac{|\Sigma_\phi|^{-1/2}}{(2\pi)^{\frac{D+2}{2}}} \times \exp\left(-\frac{1}{2}(\mathbf{x}^n - \mathbf{x}_\phi^i)^T \left(\Sigma_\phi\right)^{-1} (\mathbf{x}^n - \mathbf{x}_\phi^i)\right), \quad (2)$$

where $\{\mathbf{x}_\beta^i\}_{i=1}^{N_\beta}$ is the set of N_β background reference samples, $\{\mathbf{x}_\phi^i\}_{i=1}^{N_\phi}$ represents the set of N_ϕ foreground reference samples, Σ_β and Σ_ϕ are symmetric positivedefinite $(D+2) \times (D+2)$ bandwidth matrices [5], γ is the constant density of a uniform random variable in the set of $D+2$ components, and α is a mixture factor.

Common detection strategies [18] assume that the expressions $M_\beta(\mathbf{x}^n)$ and $M_\phi(\mathbf{x}^n)$ represent directly the joint spatioappearance distributions for background, $p(\mathbf{a}^n, \mathbf{s}^n|\beta)$, and foreground, $p(\mathbf{a}^n, \mathbf{s}^n|\phi)$, respectively. This assumption leads to the computation of the probability of p^n to belong to the foreground class, via Bayes' theorem, as

$$\Pr(\phi|\mathbf{a}^n, \mathbf{s}^n) = \frac{\Pr(\phi) M_\phi(\mathbf{a}^n, \mathbf{s}^n)}{\Pr(\beta) M_\beta(\mathbf{a}^n, \mathbf{s}^n) + \Pr(\phi) M_\phi(\mathbf{a}^n, \mathbf{s}^n)}, \quad (3)$$

where $\Pr(\phi)$ and $\Pr(\beta) = 1 - \Pr(\phi)$ are the prior probabilities of an observation in any spatial position to belong to the foreground or to the background. In addition, if no further information is available, the prior probabilities in Eq. (3) are assumed equal ($\Pr(\beta) = \Pr(\phi) = 1/2$) over the whole image extent.

The previously defined Bayesian classifier does not allow the integration of additional sources of prior information depending on the spatial position of each pixel. To overcome this important limitation, we propose a novel and efficient adaptable Bayesian classifier. Our proposal, inspired by an elegant and useful reinterpretation of the classical Bayesian classifier, makes it possible to inject additional spatial prior information into the system, while retaining the discriminative capabilities of the nonparametric appearance models at each pixel position. For this purpose, we stem now from the alternative decomposition,

$$\Pr(\phi|\mathbf{a}^n, \mathbf{s}^n) p(\mathbf{a}^n|\mathbf{s}^n) = p(\phi, \mathbf{a}^n|\mathbf{s}^n) = \Pr(\phi|\mathbf{s}^n) p(\mathbf{a}^n|\mathbf{s}^n, \phi), \quad (4)$$

to obtain the equivalent expression

$$\Pr(\phi|\mathbf{a}^n, \mathbf{s}^n) = \frac{\Pr(\phi|\mathbf{s}^n) p(\mathbf{a}^n|\mathbf{s}^n, \phi)}{\Pr(\beta|\mathbf{s}^n) p(\mathbf{a}^n|\mathbf{s}^n, \beta) + \Pr(\phi|\mathbf{s}^n) p(\mathbf{a}^n|\mathbf{s}^n, \phi)}, \quad (5)$$

where $\Pr(\phi|\mathbf{s}^n)$ and $\Pr(\beta|\mathbf{s}^n)$ are prior probabilities defined for each specific position $\mathbf{s}^n$, which can be expressed as

$$\Pr(\phi|\mathbf{s}^n) = \frac{\Pr(\phi) p(\mathbf{s}^n|\phi)}{\Pr(\beta) p(\mathbf{s}^n|\beta) + \Pr(\phi) p(\mathbf{s}^n|\phi)} \quad (6)$$

and $\Pr(\beta|\mathbf{s}^n) = 1 - \Pr(\phi|\mathbf{s}^n)$.

This rewriting allows a more comprehensive interpretation of the inherent prior spatial distributions assumed by the classical classifier, as illustrated in Fig. 1. The left part of this figure shows the decomposition of the spatioappearance background model into its marginal distribution of $\mathbf{s}^n$ and its conditional appearance distribution. In the opposite part of the figure, the corresponding decomposition of the spatioappearance foreground model is depicted. Finally, the figure illustrates how the decomposed versions of both original distributions are combined with the background and foreground prior probabilities in a Bayesian classifier.

According to this interpretation of the classical classifier, the joint spatioappearance distributions $p(\mathbf{a}^n, \mathbf{s}^n|\beta)$ and $p(\mathbf{a}^n, \mathbf{s}^n|\phi)$ are respectively decomposed into the product of two factors: (1) the marginal distributions of $\mathbf{s}^n$, $p(\mathbf{s}^n|\beta)$ and $p(\mathbf{s}^n|\phi)$, which in the conditions of Eq. (3) yield

$$p(\mathbf{s}^n|\beta) = M_\beta(\mathbf{s}^n) = \int M_\beta(\mathbf{a}^n, \mathbf{s}^n) \mathrm{d}\mathbf{a}^n, \quad (7)$$

$$p(\mathbf{s}^n|\phi) = M_\phi(\mathbf{s}^n) = \int M_\phi(\mathbf{a}^n, \mathbf{s}^n) \mathrm{d}\mathbf{a}^n, \quad (8)$$

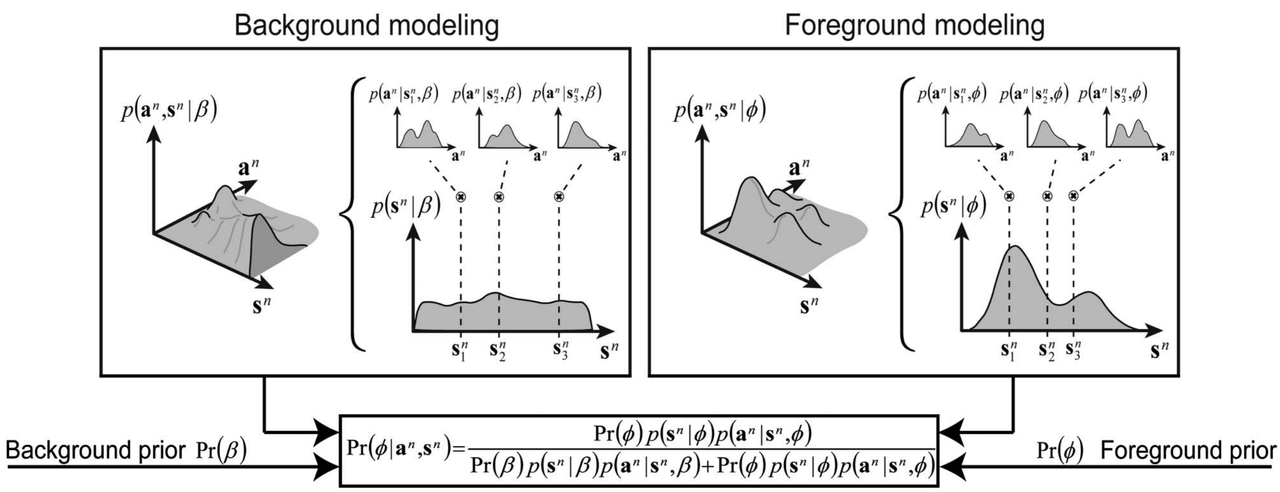

Fig. 1. Bayesian classifier from the combination of spatioappearance background and foreground estimated models decomposed in: spatial marginal distributions, conditional appearance distributions, and prior probabilities. To facilitate the graphical representations only one appearance component and one spatial component have been considered.

and (2) the conditional appearance distributions $p(\mathbf{a}^n|\mathbf{s}^n, \beta)$ and $p(\mathbf{a}^n|\mathbf{s}^n, \phi)$ for each spatial position, which in the discussed case are

$$p(\mathbf{a}^n|\mathbf{s}^n, \beta) = \frac{M_\beta(\mathbf{a}^n, \mathbf{s}^n)}{M_\beta(\mathbf{s}^n)}, \quad (9)$$

$$p(\mathbf{a}^n|\mathbf{s}^n, \phi) = \frac{M_\phi(\mathbf{a}^n, \mathbf{s}^n)}{M_\phi(\mathbf{s}^n)}, \quad (10)$$

for the background and for the foreground, respectively. Therefore, although the classical classifier described in Eq. (3) does not apparently inject any prior information on foreground object locations in the considered frame, it does not only inject appearance models for each spatial position through the conditional appearance distribution defined in Eqs. (9) and (10), but also forces implicitly a spatial prior probability function, $\Pr(\phi|\mathbf{s}^n)$, directly determined by the sample sets used for appearance modeling through

$$\Pr(\phi|\mathbf{s}^n) = \frac{\Pr(\phi)M_\phi(\mathbf{s}^n)}{\Pr(\beta)M_\beta(\mathbf{s}^n) + \Pr(\phi)M_\phi(\mathbf{s}^n)}. \quad (11)$$

Common detection strategies sticking to Eq. (3) are thus unable to integrate naturally multiple sources of spatial prior information. To overcome this limitation, we retake Eq. (5) to propose a flexible classifier that retains the appearance modeling performed for each individual spatial position $\mathbf{s}^n$ by maintaining the conditional distributions given in Eqs. (9) and (10), but allowing the integration of additional sources of spatial information through the definition of the modified prior spatial distribution

$$\tilde{\Pr}(\phi|\mathbf{s}^n) \equiv \frac{N_\phi(\mathbf{s}^n)M_\phi(\mathbf{s}^n)}{N_\beta(\mathbf{s}^n)M_\beta(\mathbf{s}^n) + N_\phi(\mathbf{s}^n)M_\phi(\mathbf{s}^n)}, \quad (12)$$

where $M_\phi(\mathbf{s}^n)$ and $M_\beta(\mathbf{s}^n)$ are the spatial marginal versions of the nonparametric models, as defined in Eqs. (7) and (8), and $N_\phi(\mathbf{s}^n)$ and $N_\beta(\mathbf{s}^n)$ are any pair of strictly positive functions whose quotient represents the relative prior confidence that the spatial position $\mathbf{s}^n$ is occupied by a moving object, according to all the additional available sources on object presence at time n. The definition in Eq. (12) is a perfectly valid spatial prior function, whose main advantage is to allow the satisfactory fusion of multiple information sources, including naturally the observed samples used for appearance modeling, but not restricting prior information to them.

This beneficial characteristic of the proposed Bayesian classifier, that is, the chance to use prior spatial functions obtained from any source of information, makes it perfectly suitable not only for the moving object detection but also for many other computer vision and optics applications using similar Bayesian methods. Some good examples of these are: anomaly detection in infrared and hyperspectral imaging, detection and characterization of discrete objects in astrophysics and cosmology, or salient object detection in static images.

We have presented a novel, efficient, and adaptable Bayesian classifier that allows the natural combination of nonparametric appearance modeling of foreground and background objects with additional sources of prior information on moving object detection location. The definition of this prior information is carried out in a practical and convenient way, expressed as a combination of strictly positive spatial functions defined over the spatial domain of the images.

Although this classifier has been designed for moving object detection strategies, its use is not limited to this type of task. Its ability to include any additional prior information obtained from unspecific sources of information (depending on the spatial position of each pixel) makes it perfectly suitable for many others fields of research (e.g., computer vision and optics applications), where Bayesian classifiers are required.

References

1. K. Sangani, Eng. Technol. **5**, 28 (2010).
2. Y. Sheikh, O. Javed, and T. Kanade, in *Proceedings of IEEE 12th International Conference on Computer Vision* (IEEE, 2009), pp. 1219–1225.
3. B. Zhang, Y. Gao, S. Zhao, and B. Zhong, IEEE Trans. Circuits Syst. Video Technol. **21**, 29 (2011).
4. M. Cristani, M. Farenzena, D. Bloisi, and V. Murino, EURASIP J. Adv. Sig. Proc. **2010**, 343057 (2010).
5. A. Tavakkoli, M. Nicolescu, G. Bebis, and M. Nicolescu, Machine Vis. Appl. **20**, 395 (2009).
6. S. Y. Elhabian, K. M. El-Sayed, and S. H. Ahmed, Rec. Pat. Comput. Sci. **1**, 32 (2008).
7. A. Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis, Proc. IEEE **90**, 1151 (2002).
8. T. Tanaka, A. Shimada, R. Taniguchi, T. Yamashita, and D. Arita, Comput. Vis. Image Underst. **113**, 63 (2009).
9. Y. Sheikh and M. Shah, IEEE Trans. Pattern Anal. Mach. Intell. **27**, 1778 (2005).
10. X. Zhang and J. Yang, Electron. Lett. **44**, 851 (2008).
11. C. Cuevas, D. Berjón, F. Morán, and N. García, IEEE Trans. Consum. Electron. **58**, 117 (2012).
12. Q. Baig, T. D. Vu, and O. Aycard, in *Proceedings of IEEE International Conference on Intelligent Computer Communication and Processing* (IEEE, 2009), pp. 401–408.
13. H. Yang, J. Tian, Y. Chu, Q. Tang, and J. Liu, IEEE Signal Process. Lett. **15**, 497 (2008).
14. F. A. Sadjadi and C. S. L. Chun, Opt. Lett. **28**. 531 (2003).
15. R. S. DiPietro, D. Manolakis, R. Lockwood, T. Cooley, and J. Jacobson, Proc. SPIE **7695**, 1 (2010).
16. P. Carvalho, G. Rocha, and M. P. Hobson, Mon. Not. R. Astron. Soc. **393**, 681 (2009).
17. W. Zhang, Q. M. J. Wu, G. H. Wang, and H. B. Yin, IEEE Trans. Multimedia **12**, 300 (2010).
18. N. Martel-Brison and A. Zaccarin, in *Proceedings of the 1st ACM Workshop on Vision Networks for Behavior Analysis* (ACM, 2008), pp. 93–100.