Image and Vision Computing 31 (2013) 616–630

Contents lists available at SciVerse ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Improved background modeling for real-time spatio-temporal non-parametric moving object detection strategies ☆

Carlos Cuevas *, Narciso García

Grupo de Tratamiento de Imágenes, Universidad Politécnica de Madrid (UPM), E-28040 Madrid, Spain

ARTICLE INFO

Article history:
Received 7 September 2012
Received in revised form 11 February 2013
Accepted 3 June 2013

Keywords:
Dynamic bandwidth estimation
Moving object detection
Non-parametric background modeling
Selective update
Spatio-temporal data

ABSTRACT

Answering to the growing demand of machine vision applications for the latest generation of electronic devices endowed with camera platforms, several moving object detection strategies have been proposed in recent years. Among them, spatio-temporal based non-parametric methods have recently drawn the attention of many researchers. These methods, by combining a background model and a foreground model, achieve high-quality detections in sequences recorded with non-completely static cameras and in scenarios containing complex backgrounds. However, since they have very high memory and computational associated costs, they apply some simplifications in the background modeling process, therefore decreasing the quality of the modeling.

Here, we propose a novel background modeling that is applicable to any spatio-temporal non-parametric moving object detection strategy. Through an efficient and robust method to dynamically estimate the bandwidth of the kernels used in the modeling, both the usability and the quality of previous approaches are improved. Furthermore, by adding a novel mechanism to selectively update the background model, the number of misdetections is significantly reduced, achieving an additional quality improvement. Empirical studies on a wide variety of video sequences demonstrate that the proposed background modeling significantly improves the quality of previous strategies while maintaining the computational requirements of the detection process.

1. Introduction

The recent significant growth of electronic devices endowed with video cameras [1] has resulted in an important demand for new machine vision application tools [2]. To perform high-level tasks (e.g., tracking, classification, event analysis, or augmented reality) these tools include, as a key step, a moving object detection strategy [3]. Owing to the importance of these strategies, several approaches have been proposed [4] to efficiently detect moving objects.

Some algorithms aim to reduce the memory requirements and to maximize the speed [5], providing successful results for short sequences with simple backgrounds [6]. However, they do not provide satisfactory results in complex scenarios with illumination changes, shadows, or dynamic backgrounds [7].

To improve the quality of the detections in complex scenarios, several multimodal strategies have been proposed [8]. One key multimodal method is the Mixture of Gaussians (MoGs) model proposed by Stauffer and Grimson [9], which makes use of a mixture of several Gaussians to obtain an adaptive model of each image pixel [10]. Other popular multimodal methods are those using Hidden Markov Models (HMMs) [11], which model the background variations by representing the changes in the scene with different states (day or night, lights on or lights off, etc.).

Although these methods are able to provide high-quality detections in many scenarios, they fail in environments where the pixel statistics cannot be described parametrically [12]. To improve the quality of the results in these environments, several moving object detection strategies using non-parametric kernel density estimation methods have been developed [13]. Instead of considering the values of the pixels as a particular distribution, these methods obtain probabilistic models from sets of recent reference samples [14].

Among non-parametric strategies, a significant number of spatio-temporal background–foreground modeling approaches [15] have been recently proposed to improve the quality of the detections in sequences recorded with non-completely static cameras or containing non-static background regions [16]. These approaches construct both models (background and foreground) by using spatio-temporal reference data [17]. In this way, they improve the quality of the detections provided by other non-parametric methods [18]. However, since these strategies have associated very high memory and computational costs [19], they carry out some simplifications that reduce the quality of the background modeling. They model the background with fixed bandwidth matrices [16], decreasing the quality of the estimations and forcing to manually set appropriate bandwidth values

☆ This paper has been recommended for acceptance by Bodo Rosenhahn.

* Corresponding author. Tel.: +34 913367353.
E-mail addresses: ccr@gti.ssr.upm.es (C. Cuevas), narciso@gti.ssr.upm.es (N. García).

0262-8856/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.imavis.2013.06.003

depending on the characteristics of each sequence. Furthermore, instead of applying the selective mechanisms used by other strategies to update the background model, they use blind mechanisms [14], which significantly increase the amount of misdetections.

In this paper, we propose a real-time high-quality background modeling that is suitable for integration into any moving object detection strategy based on spatio-temporal non-parametric modeling. We robustly estimate the bandwidth of the kernels from spatially weighted distributions of the differences of reference samples for consecutive images. Additionally, we apply a novel mechanism to efficiently update the background model, which applies to each reference sample a weight factor determining its influence in the modeling.

The rest of the paper is organized as follows. Section 2 presents the state of the art corresponding to the most relevant work in kernel bandwidth estimation (Section 2.1) and background update (Section 2.2). Section 3 introduces the notation corresponding to the moving object detection strategies based on non-parametric modeling. Section 4 describes the method proposed to dynamically estimate the bandwidth of the kernels. Section 5 presents the algorithm proposed to update the background model. Finally, Sections 6 and 7 present, respectively, the obtained results and the conclusions.

2. Related work

2.1. Kernel bandwidth estimation

Within non-parametric moving object detection strategies, the selection of adequate bandwidth matrices for the used kernels is crucial to obtain high quality results [20]. Using too small bandwidth values, the amount of false detections increases while, if too large bandwidth values are used, the multimodality can be lost. The optimal bandwidth matrix can be theoretically found by minimizing the mean-squared error between the estimated density function, $\hat{f}$, and the true density, f. However, as this error depends on the unknown true density function, several heuristic strategies to dynamically estimate these matrices have been proposed [21].

In general, two formulations can be found in the literature [16]. The first one is known as balloon estimator and varies the bandwidth with the estimation point. This formulation is given by

$$\hat{f}(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N}\left|\sum(\mathbf{x})\right|^{-\frac{1}{2}}K\left(\left(\sum(\mathbf{x})^{-\frac{1}{2}}\left(\mathbf{x}-\mathbf{x}^{i}\right)\right)\right), \tag{1}$$

where $\sum(\mathbf{x})$ is the bandwidth matrix at the estimation point $\mathbf{x}$, $\{\mathbf{x}^i\}_{i=1}^{N}$ is a set of N samples from f, and K is a kernel function. The second one, which is known as sample point estimator, varies the bandwidth depending on the sample point and is given by

$$\hat{f}(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N}\left|\sum\left(\mathbf{x}^{i}\right)\right|^{-\frac{1}{2}}K\left(\left(\sum\left(\mathbf{x}^{i}\right)^{-\frac{1}{2}}\left(\mathbf{x}-\mathbf{x}^{i}\right)\right)\right), \tag{2}$$

where $\sum(\mathbf{x}^i)$ is the bandwidth matrix at the sample point $\mathbf{x}^i$.

Seeking simplicity and computational efficiency, some authors propose algorithms that use kernels with constant and global bandwidth values [22], which are able to provide good quality detections in sequences with slow background changes. However, they fail in scenarios containing backgrounds with complex or irregular densities [21].

To obtain high quality results, some strategies [7] make use of generic (non-necessarily diagonal) bandwidth matrices that are dynamically estimated at each pixel location. To construct these matrices, these approaches require to store large amounts of data and perform too many operations [14]. So, their associated computational and memory costs are very high.

Looking for a reasonable compromise between efficiency and quality, other strategies assume independency between all the characteristics used to represent the image pixels [23]. So, instead of using generic matrices, they use diagonal bandwidth matrices [24].

Regarding spatio-temporal non-parametric strategies, since no appropriate dynamic estimation scheme has been proposed for this kind of algorithms, they commonly use diagonal bandwidth matrices with fixed and global values [17]. In this way, they suppress the computational effort associated to the estimation of appropriate bandwidth matrices. However, the quality of the modeling decreases and, additionally, they require the manual setting of different bandwidth values depending on the characteristics of the sequences, which decreases their usability.

2.2. Background update

To adapt changes in the scene, non-parametric moving object detection strategies need to continuously update the sets of reference samples that are used to model the background [25]. Several mechanisms to efficiently update the background have been proposed in the literature. According to [26] they can be classified into: selective update mechanisms, which discard the samples from pixels that have been previously classified as foreground; and blind update mechanisms, which use all the samples extracted from the reference pixels.

Blind update mechanisms are easier to use, since they do not require the inclusion of additional strategies to decide if a pixel belongs to the background or to the foreground [27], which are usually complex and increase the computational cost of the modeling process. Consequently, blind update mechanisms have been used in several moving object detection strategies in recent years [12]. However, since these mechanisms include all the previously detected moving objects as part of the background model, they lead to significant amounts of misdetections.

Selective update mechanisms try to discard from the set of reference samples those samples previously classified as foreground [28]. In this way, as samples from foreground pixels are not used to estimate the background model, these mechanisms obtain final detections with less misdetections [29]. Currently, a large amount of moving object detection strategies proposing different selective update alternatives to model the background can be found in the literature [30]. However, as they require additional and complex object-level stages, they involve very high computational efforts [31] and, additionally, most of them cannot easily be included in any other moving object detection strategy.

Among selective update strategies, it is interesting to highlight that proposed in [32], since it has been extensively referred by many authors [26] and has been incorporated in several moving object detection strategies in recent years [33,34]. The update mechanism presented in [32] proposes to combine a short term model (a very recent model of the scene) with a long-term model (a more stable representation of the scene). The first model adapts to the fast changes in the background, while the second model adapts to its slow variations. By combining both models the quality of the background modeling is noticeably improved. However, this method has some important drawbacks. The length of both short-term and long-term models should be manually set accordingly to the content of the analyzed sequences (slow or fast moving objects, dynamic or stationary background variations, etc.). So, its usability is decreased. Additionally, as the estimation of two background models for each image is required, it involves very high memory and computational costs [35]. Therefore, this selective updating is not appropriate for spatio-temporal moving object detection strategies, since the main problem in this kind of strategies is the extremely high computational and memory costs to model the background [19]. Moreover, in case of sudden changes in lighting, significant amounts of pixels remain misclassified as foreground until the long term model adapts to the new illumination conditions [36]. So, this strategy is not appropriate for sequences with frequent changes in lighting.