Region-based Moving Object Detection Using Spatially Conditioned Nonparametric Models in a GPU

Daniel Berjón, Carlos Cuevas, Francisco Morán, and Narciso García

***Abstract*—A novel GPU-based nonparametric moving object detection strategy for computer vision tools requiring real-time processing is proposed. An alternative and efficient Bayesian classifier to combine nonparametric background and foreground models allows increasing correct detections while avoiding false detections. Additionally, an efficient region of interest analysis significantly reduces the computational cost of the detections.**

I. Introduction

The demand for commercial computer vision tools (mixed or augmented reality, tracking, event detection, etc.) for electronic devices with smart cameras is increasing rapidly [1]. Many of these tools require friendly and lightweight moving object detection strategies that must be able to provide high-quality results in complex sequences recorded with non-completely static cameras [2].

In [3] we proposed a very friendly, high-quality, and real-time GPU-based implementation of a spatio-temporal nonparametric strategy to detect moving objects. However, since the foreground modeling proposed in this strategy is uniform, it may lead to misdetections if moving objects contain regions with similar appearance to background areas (see Fig. 1.b). This drawback can be solved by combining the background model in [3] with the foreground model in [4], but then two new issues appear. Since the foreground modeling requires kernels with larger spatial widths [4], the spatial density of the foreground model is much higher than the spatial density of the background, which may lead to false detections (see Fig. 1.c). Additionally, the computational cost increases, making it difficult to integrate the detection strategy in computer vision tools requiring real-time processing.

We propose a novel GPU-based strategy that combines the abovementioned background and foreground models, while maintaining the computational requirements and avoiding the previously described false detections.

II. Spatially conditioned nonparametric modeling

Let p^n be a pixel in the current image, at time n, defined by $\mathbf{x}^n=((\mathbf{a}^n)^T,(\mathbf{s}^n)^T)^T$, where $\mathbf{a}^n$ is a D-dimensional vector containing appearance information of p^n and $\mathbf{s}^n=(h^n,w^n)$ is a vector containing its spatial coordinates (rows and columns).

For the sake of computational efficiency, spatio-temporal approaches use small spatial widths to model the background and large spatial widths to model the foreground [5].

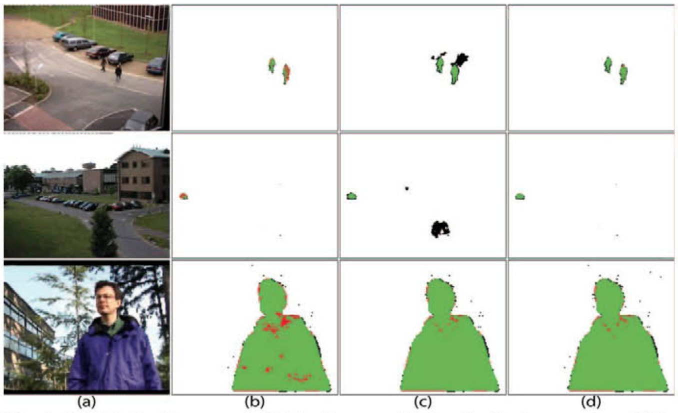

Fig. 1. (a) Original images. (b) Detections combining the background model in [3] and a uniform foreground model. (c) Detections combining the background model in [3] and the foreground model in [4] with the typical Bayesian classifier. (d) Detections combining the same models used in (c) with the proposed classifier. Color notation: correct detections (green), misdetections (red), and false detections (black).

Consequently, in some image regions, the spatial density of the foreground model can become much larger than that of the background, which can lead to significant persistent false detections if the typical Bayesian classifier [6] is used to combine both models. To avoid this important drawback we propose to use an alternative classifier defined as

$$\Pr\left(\phi \mid \mathbf{x}^n\right) = \frac{\Pr\left(\phi\right)\mathrm{p}\left(\mathbf{x}^n \mid \phi, h^n, w^n\right)}{\Pr\left(\phi\right)\mathrm{p}\left(\mathbf{x}^n \mid \phi, h^n, w^n\right) + \Pr\left(\beta\right)\mathrm{p}\left(\mathbf{x}^n \mid \beta, h^n, w^n\right)} ,$$

where $\Pr(\phi)$ and $\Pr(\beta) = \Pr(\phi) - 1$ are the prior probabilities of an observation to belong to the foreground, ϕ, or to the background, β; and $\mathrm{p}(\mathbf{x}^n \mid \phi, h^n, w^n)$ and $\mathrm{p}(\mathbf{x}^n \mid \beta, h^n, w^n)$ result from conditioning the foreground model, $\mathrm{p}(\mathbf{x}^n \mid \phi)$, and background model, $\mathrm{p}(\mathbf{x}^n \mid \beta)$, on a particular spatial location. These density functions are obtained as

$$\mathrm{p}\left(\mathbf{x}^n \mid \xi, h^n, w^n\right) = \frac{\mathrm{p}\left(\mathbf{x}^n \mid \xi\right)}{\mathrm{p}\left(h^n, w^n \mid \xi\right)} \text{, such that } \xi = \left\{\beta, \phi\right\},$$

where $\mathrm{p}(h^n, w^n \mid \xi)$ is the marginalization of $\mathrm{p}(\mathbf{x}^n \mid \xi)$ over the D-dimensional set of appearance characteristics.

Figure 1 illustrates some final results in three different scenarios. Fig. 1.b shows the results obtained by combining the background modeling in [3] and a uniform foreground model. Note that using this combination several misdetections are obtained. Fig. 1.c and Fig. 1.d present the detections resulting from the combination of the background modeling in [3] and the foreground modeling in [4], using the typical Bayesian classifier [6] (Fig. 1.c), and using the one we propose (Fig. 1.d). It can be observed that misdetections are avoided by adding the non-uniform foreground model. However, it is only thanks to the proposed alternative classifier that false detections are also avoided.

III. REGION-BASED ANALYSIS

In order to reduce computation time, we propose a selective analysis of the image based on Region of Interest (RoI) masks. We assume that a meaningful moving object needs to be larger than a (configurable) given size and that the temporal sampling rate of the sequence is high enough so that there is some overlap between the regions occupied by a given object in two consecutive frames.

For each new frame we focus our analysis on a small number of tiles randomly placed within a grid (shown in solid orange in Fig. 2.c) to detect new objects, and the tiles where a moving object was detected in the previous frame (translucent orange in Fig. 2.b and 2.c); then, we check which of these tiles actually result in a positive detection, examine its adjacent tiles (translucent pink in Fig. 2.c), and iterate this process until no more tiles need to be analyzed.

This strategy effectively reduces the amount of data that must be accessed, but breaks its regularity: it is no longer adequate to map a 2D array of threading blocks in the GPU to the 2D array of image tiles because most blocks would have no work to perform. Instead, we use a 1D list containing the indices of the tiles to be worked on by the GPU.

To facilitate the generation of this work list, we use three auxiliary thumbnail images (each pixel representing a tile of the original frame) to keep track of which tiles need to be processed, have already been processed, or have been found to contain a foreground object. After each analysis pass, the thumbnail representing positive detections undergoes morphological dilation to generate new tiles to be analyzed, then those blocks that have not already been analyzed are inserted into the next work list for the GPU. Since the thumbnails are very small and the work list must be coherent, this bookkeeping task is carried out in the CPU.

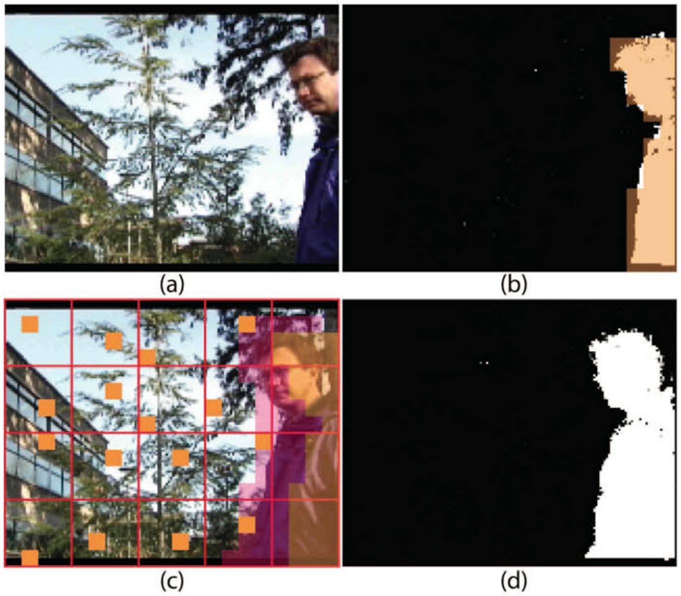

Fig. 2. (a) Image I^n. (b) Detection obtained for I^n (in white) and mask of blocks to analyze in the next image (in orange). (c) Image I^{n+1}, random blocks to analyze (in orange), mask of blocks obtained from the detection in I^n, and additional blocks analyzed (in pink). (d) Detection for I^{n+1} (in white).

IV. RESULTS

Tables II and III show some results obtained for the sequences described in Table I. It can be observed that using the proposed RoI masks the overall computational cost is significantly reduced. This reduction depends on the percentage of foreground pixels along the sequences: larger percentages result in larger RoI masks.

TABLE I
DESCRIPTION OF THE TEST SEQUENCES

Sequence	Size (H×W)	No. of images	No. of moving objects	Foreground pixels
S1(W2)	128×160	287	1	1.52%
S2(L3)	192×256	550	1	2.72%
S3(P1)	288×352	1452	8	1.12%

TABLE II
MEAN PROCESSING TIMES PER FRAME WITHOUT ROI MASKS

Sequence	Background modeling	Foreground modeling	Total
S1	6 ms	17 ms	23 ms
S2	13 ms	37 ms	50 ms
S3	30 ms	81 ms	111 ms

TABLE III
MEAN PROCESSING TIMES PER FRAME WITH ROI MASKS

Sequence	Computed pixels	Background modeling	Foreground modeling	Total
S1	10.39%	0.8 ms	2.3 ms	3.1 ms
S2	17.46%	3.0 ms	10.2 ms	13.2 ms
S3	9.68%	3.8 ms	10.8 ms	14.6 ms

V. CONCLUSIONS

A GPU-based nonparametric moving object detection strategy has been proposed, which is suitable for computer vision tools demanded by the last generation of consumer electronic devices, where both high-quality results and real-time processing are required. Using an alternative Bayesian classifier, background and foreground models are combined while avoiding most false detections. Additionally, through a selective analysis of the RoIs in the images the computational requirements of the detection process in maintained.

REFERENCES

[1] G. Shapiro, "Consumer Electronics Association's Five Technology Trends to Watch: Exploring New Tech That Will Impact Our Lives", IEEE Consumer Electronics Magazine, vol. 2, no. 1, pp. 32-35, 2013.

[2] W. La, J. Han, and P. H. N. de With, "Automatic video-based human motion analyzer for consumer surveillance system," *IEEE Trans. Consumer Electronics*, vol. 55, no. 2, pp. 591-598, 2009.

[3] D. Berjón, C. Cuevas, F. Morán, N. García, "GPU-based Implementation of an Optimized Nonparametric Background Modeling for Real-time Moving Object Detection", *IEEE Trans. Consumer Electronics*, vol. 59, no. 2, pp. 361-369, 2013.

[4] C. Cuevas, D. Berjón, F. Morán, N. García, "Moving Object Detection for Real-Time Augmented *Reality Applications in a GPGPU", IEEE* Trans. Consumer Electronics, vol. 58, no. 1, pp. 117-125, 2012.

[5] C. Cuevas, N. García, "Efficient Moving Object Detection for Lightweight Applications on Smart Cameras", *IEEE Trans. Circuits and Systems for Video Technology*, vol. 23, no. 1, pp. 1-14, 2013

[6] Y. Sheikh and M. Shah, "Bayesian modeling of dynamic scenes for object detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 11, pp. 1778-1792, 2005.