Statistical Moving Object Detection for Mobile Devices with Camera

Carlos Cuevas, Raúl Mohedano, and Narciso García

***Abstract*—A novel and high-quality system for moving object detection in sequences recorded with moving cameras is proposed. This system is based on the collaboration between an automatic homography estimation module for image alignment, and a robust moving object detection using an efficient spatio-temporal nonparametric background modeling.**

I. Introduction

The recent proliferation of electronic devices with a camera has greatly increased the demand for computer vision applications [1]. Many of these applications need to include high-quality moving object detection methods as a first step for higher level analysis tasks (*e.g.* tracking, mixed or augmented reality). Among the many methods proposed in the last decades to detect moving objects [2], two main general statistical modeling background approaches must be outlined: mixture-of-Gaussians (MoG), based on parametric methods [3], and nonparametric modeling, using kernel density estimation [4]. The growing computational power of current processors is causing an increasing preference for the latter, since it shows better adaptation capabilities and more natural handling of spatial uncertainties caused by camera instability and dynamic backgrounds [5].

Most parametric and nonparametric algorithms have been designed for static cameras [6]. However, there is a significant need for efficient strategies for moving object detection in sequences recorded with moving camera platforms (*e.g.* cellular phones, vehicles, or robots), able to handle apparent displacements due to sensor movement and not to the presence of moving agents.

We propose a novel and efficient system for moving object detection in sequences recorded with non-static cameras. This system automatically estimates the homography compensating the apparent scene background motion induced by the moving camera, and creates a spatio-temporal kernel-based background model from reference pixel observations from previous images. adapted spatially using the inferred transforms. Robust homography estimation is necessary to spatially align each current sample with its corresponding reference data, while spatio-temporal modeling is required to avoid the inevitable inaccuracies in data alignment. In addition, final detection results are fed back into the estimation of subsequent homographies, reducing thus the influence of foreground features and improving the accuracy of the resulting alignment.

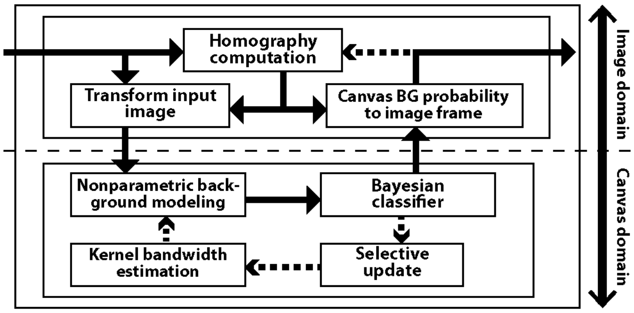

Fig. 1. Flowchart of the proposed strategy.

II. Proposed Strategy

Let I^n be the n-th frame of a video sequence, composed of a set of pixels defined as (D+2)-dimensional vectors $\mathbf{x}^n = (\mathbf{a}^n, \mathbf{s}^n)$, where $\mathbf{a}^n \in \mathbb{R}^D$ represents the color value of the pixel and $\mathbf{s}^n = (u^n, v^n) \in \mathbb{R}^2$ its spatial coordinates within the coordinate system $\mathfrak{R}^n$ of the image. Let such video sequence be captured by a camera undergoing a movement not causing visual parallax in the static parts of the background (*e.g.* PTZ moves, displacements of negligible amplitude compared to the distance to the background).

Under these assumptions, the apparent displacement of the static parts of the background can be ideally compensated using a projective transform: every pixel $\mathbf{x}^n$ in I^n corresponding to a 3D point of the the static background can be related, using the appropriate homography $h_{n-1}^n(\mathbf{s}^n)$, to the position of the coordinate system $\mathfrak{R}^{n-1}$ of I^{n-1} where that 3D point would be projected. The theoretical composition of successive homographies shows that, ideally, there is a homography transforming each pixel position $\mathbf{s}^n$ in I^n (belonging to the scene background) into the position in $\mathfrak{R}^m$ $(1 \leq m < n)$ where the corresponding 3D point would be observed if the camera setting were the same as in frame m but its FoV were unlimited.

The previous reasoning justifies the establishment of a unified coordinate system $\mathfrak{R}^E$, corresponding to one of the previously processed video frames (arbitrarily chosen since the properties of the homographies allows the use of any of them), to set a virtual canvas where the spatio-temporal nonparametric background modeling is performed. With this approach, if the corresponding background homography h_E^n for each incoming video frame can be estimated, the pixels composing the image I^n can be transformed into $\mathfrak{R}^E$ (from image to canvas domain, see Fig. 1): the transformed image $\tilde{I}_E^n$, defined as

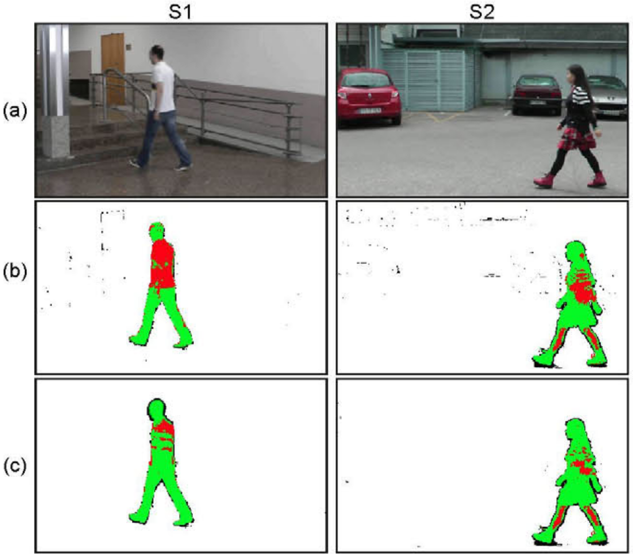

Fig. 2. Original images (a) and results using: a spatio-temporal nonparametric background modeling (b) and the proposed strategy (c). Color notation: correct detections (green), misdetections (red), and false detections (black).

the set of spatially-adapted pixels $\tilde{\mathbf{x}}_E^n = (\mathbf{a}^n, h_E^n(\mathbf{s}^n))$, is then incorporated to the set of reference samples used in $\mathfrak{R}^E$ to create the nonparametric model in the virtual canvas.

The probabilistic foreground/background classification of the pixels of such virtual canvas is also the basis for the classification of each original input pixel $\mathbf{x}^n$, obtained by transforming the probabilistic classification performed on the canvas and applying the inverse of the transformation h_E^n to obtain the final interpolated probability mask for I^n (from canvas to image domain, see Fig. 1).

III. Results

The proposed strategy has been tested in two sequences, S1 and S2, recorded with non static cameras and containing critical situations for moving object detection strategies. S1 is an indoor sequence captured by a camera panning from right to left. S2 is an outdoor sequence captured by a shaking camera and presenting continuous background jolts in all directions.

Figure 2.a depicts one original image from each test sequence. The detections obtained with the strategy in [7], which is an improved version of the spatio-temporal nonparametric method for non-static cameras in [6], are depicted in Fig. 2.b. The detections obtained with the proposed mosaicing-modeling feedback appear in Fig. 2.c. These results show that the proposed strategy avoids most false detections due to camera motion and that it obtains compact detected regions.

Finally, Fig. 3 shows the overall $recall$, $precision$ and F percentages [6] (evaluating jointly both $recall$ and $precision$ as $F = 2\frac{rec \times pre}{rec + pre}$) obtained with the proposed strategy and with the modeling in [7] for different spatial kernel widths (σ). As expected, in the modeling proposed in [7], the amount of false detections decreases (higher $precision$) as the used spatial bandwidth increases. However, the amount of correct detections also decreases (lower $recall$), which results in low F values. As our strategy updates the spatial position of the background reference data, we provide both high $recall$ and $precision$ percentages using a small spatial bandwidth. So, we clearly obtain the best F values.

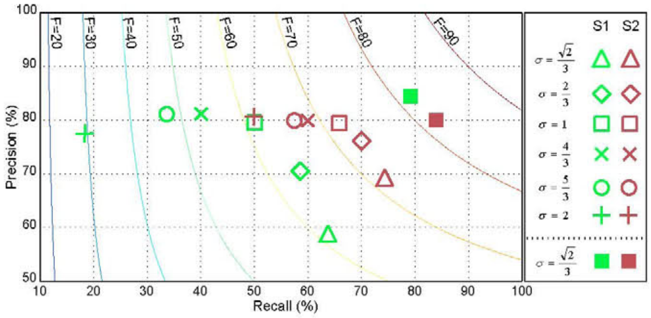

Fig. 3. Quantitative comparison between the strategy in [7] (for different spatial kernel widths σ) and our proposal (the last two filled squares). Isolines for different values of the figure of merit F are superimposed.

IV. Conclusion

A novel moving object detection algorithm for sequences recorded with non-static camera has been proposed, suitable for the computer vision applications demanded by portable device users. To this end, the motion of the scene background is compensated through automatic homography estimation. The aligned data are used as reference to model the background with a spatio-temporal nonparametric modeling, avoiding misdetections due to the inevitable inaccuracies in data alignment. The detection results are fed back to the homography module to reduce the influence of the foreground in the motion compensation. The obtained results have shown that the proposed strategy significantly improves the quality of previous approaches in the considered moving settings.

Acknowledgment

This work has been partially supported by the Ministerio de Ciencia e Innovación of the Spanish Government under projects TEC2010-20412 (Enhanced 3DTV) and TEC2013-48453 (MR-UHDTV).

References

[1] G. Shapiro, "Consumer electronics association's five technology trends to watch: Exploring new tech that will impact our lives," *Consumer Electronics Magazine, IEEE*, vol. 2, no. 1, pp. 32–35, 2013.

[2] A. Sobral and A. Vacavant, "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos," *Computer Vision and Image Understanding*, vol. 122, pp. 4–21, 2014.

[3] 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.

[4] A. Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis, "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance," *Proceedings of the IEEE*, vol. 90, no. 7, pp. 1151–1163, 2002.

[5] D. Berjon, C. Cuevas, F. Moran, and N. Garcia, "Gpu-based implementation of an optimized nonparametric background modeling for real-time moving object detection," *IEEE Trans. Consumer Electronics*, vol. 59, no. 2, 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.

[7] C. Cuevas and N. García, "Improved background modeling for real-time spatio-temporal non-parametric moving object detection strategies," *Image and Vision Computing*, vol. 31, no. 9, pp. 616–630, 2013.