

Kernel bandwidth estimation for moving object detection in non-stabilized cameras

Carlos Cuevas, Raúl Mohedano, and Narciso García
 Universidad Politécnica de Madrid, E.T.S.I.
 Telecomunicación, Grupo de Tratamiento de Imágenes,
 28040 Madrid, Spain
 E-mail: ccr@gti.ssr.upm.es

Abstract. Sophisticated strategies have been recently proposed for the detection of moving objects in non-stabilized camera setups. These strategies model both, background and foreground, using spatio-temporal non-parametric estimation. However, as no appropriate methods for dynamical kernel bandwidth are available, high-quality results cannot be obtained in all situations. Here, an automatic and efficient kernel bandwidth estimation strategy for spatio-temporal modeling is proposed. Background kernel bandwidth is estimated via a novel statistical analysis of spatially weighted data distributions, whereas foreground kernel bandwidth is estimated using a mean shift based analysis of previously detected foreground regions. © 2012 Society of Photo-Optical Instrumentation Engineers (SPIE). [DOI: 10.1117/1.OE.51.4.040501]

Subject terms: kernel bandwidth estimation; moving object detection; spatio-temporal background-foreground modeling.

Paper 111553L received Dec. 13, 2011; revised manuscript received Feb. 2, 2012; accepted for publication Feb. 13, 2012; published online Apr. 6, 2012.

1 Introduction

Moving object detection is a key step for high-level object analysis tasks such as tracking and event analysis.¹ Several non-parametric strategies for background and foreground object modeling have been proposed in the literature.² These schemes model pixel statistics using kernel density estimation, taking pixel values in previous frames as empirical reference samples. They yield high-quality detections in multimodal scenarios² and show great flexibility and adaptive capacity, and are thus, in principle, perfect for scene modeling and motion detection.

Recently, interesting extensions for these non-parametric strategies have been proposed,³ which build complex background and foreground models considering jointly color and spatial domains. The models are built using reference samples from multiple neighboring pixels in several previous frames, and are thus suitable for cameras without image stabilization.

However, these strategies depend strongly on kernel bandwidth selection,² and in addition, no appropriate dynamic estimation scheme has been proposed for this type of spatio-temporal modeling.⁴ Therefore, automatic and dynamic strategies for kernel bandwidth estimation are highly needed.⁵

In this letter, an efficient automatic bandwidth estimation strategy is presented, which is specially designed for moving

object detection based on background-foreground spatio-temporal non-parametric modeling. On the one hand, for background modeling, kernel bandwidth is derived via a robust statistical analysis of empirical color distributions, weighted according to their spatial location in the images. On the other hand, foreground bandwidth estimation is carried out by means of a novel and fast mean shift based clustering combined with a particle filter-based tracking strategy.

2 Spatio-Temporal Non-Parametric Modeling

Let us define a pixel in the current image domain as a five-dimensional vector $\mathbf{p} = (R, G, B, r, c)$, where (R, G, B) are its RGB color components and (r, c) are its spatial coordinates. The probability density function for both background, β , and foreground, φ , can be estimated from five-dimensional (spatio-color) reference samples via multidimensional kernels. The probability of each pixel to belong to the foreground class can be obtained through Bayes' theorem as:

$$\Pr(\varphi|\mathbf{p}) = \frac{\Pr(\varphi)\hat{f}(\mathbf{p}|\varphi)}{\Pr(\beta)\hat{f}(\mathbf{p}|\beta) + \Pr(\varphi)\hat{f}(\mathbf{p}|\varphi)}, \quad (1)$$

where $\hat{f}(\mathbf{p}|\beta)$ and $\hat{f}(\mathbf{p}|\varphi)$ are, respectively, the estimated background and foreground density functions, $\Pr(\varphi)$ represents the foreground prior probability, and $\Pr(\beta) = 1 - \Pr(\varphi)$ is the background prior probability.

3 Background Modeling

Let us consider a set of N_β background reference samples, $\mathbf{x}_i = (R_i, G_i, B_i, r_i, c_i)$, from the T_β previous images ($T_\beta \leq N_\beta$). Applying Gaussian kernels the background density function can be estimated as:

$$\hat{f}(\mathbf{p}|\beta) = \frac{1}{N_\beta(2\pi)^{5/2}|\mathbf{H}_\beta|^{1/2}} \times \sum_{i=1}^{N_\beta} \prod_{j=1}^5 \exp\left\{-\frac{[\mathbf{p}(j) - \mathbf{x}_i(j)]^2}{2\mathbf{H}_\beta(j,j)}\right\}, \quad (2)$$

where $\mathbf{H}_\beta = \text{diag}(h_{\beta,R}^2, h_{\beta,G}^2, h_{\beta,B}^2, h_{\beta,r}^2, h_{\beta,c}^2)$. The reference samples available for the background modeling are uniformly distributed in space. So, kernels with fixed spatial bandwidths, $h_{\beta,r} = h_{\beta,c} = h_s$, are used. However, the bandwidths corresponding to the RGB color components ($h_{\beta,R}, h_{\beta,G}, h_{\beta,B}$) should be estimated. For this purpose, we propose a robust statistical method for the estimation of the variance of the underlying color distributions at \mathbf{p} , which will be used as kernel bandwidths. Our algorithm draws on,⁶ which robustly estimates the bandwidth for each color component, $\mathbf{H}_\beta(j,j)$, through the analysis of the distribution of reference sample differences between consecutive images. Because this distribution is a Gaussian density function $N[0, 2\mathbf{H}_\beta(j,j)]$, the median of the distribution of the absolute differences, m_j , is equivalent to its quarter percentile. Therefore, the kernel bandwidth is:

$$\mathbf{H}_\beta(j,j) = \frac{m_j}{0.68\sqrt{2}}, \quad j \in \{1, 2, 3\}. \quad (3)$$

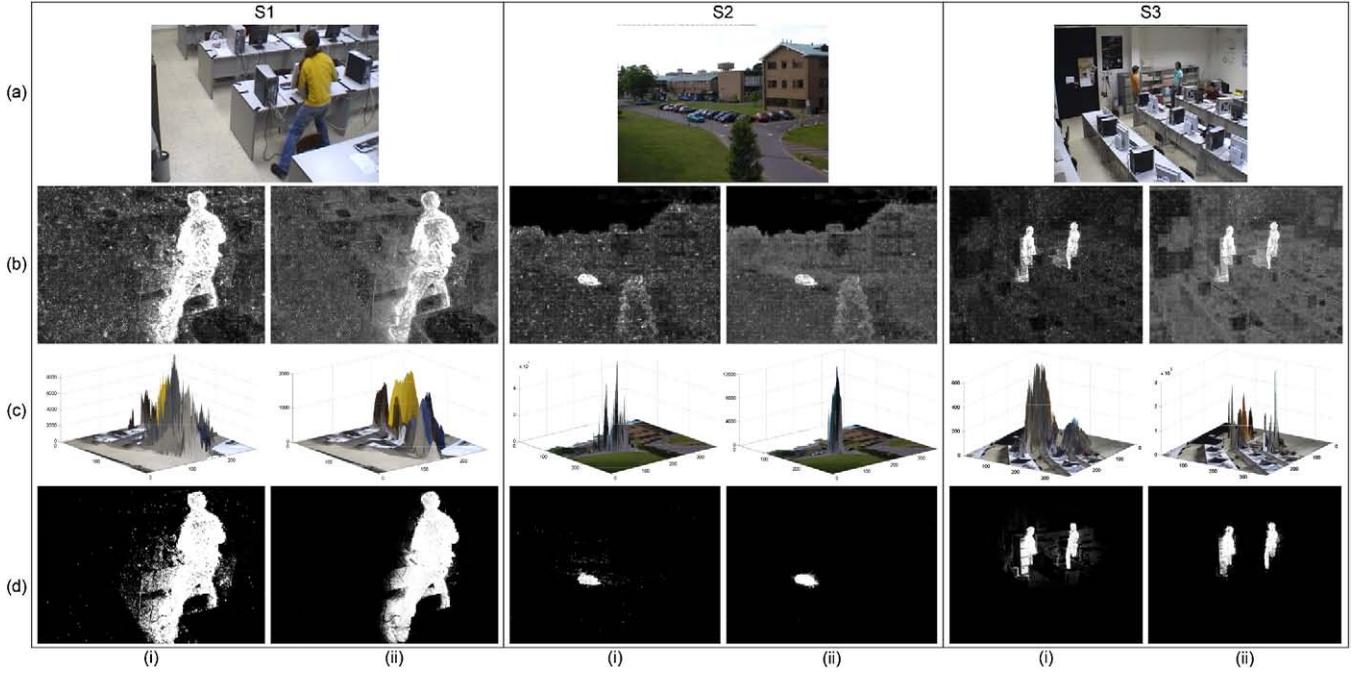


Fig. 1 (a) Original images. (b) Background log-negative likelihood. (c) Three dimensional representation of the foreground likelihood. (d) Final detections; (i) results with Ref. 3, and (ii) our results.

The proposed spatio-temporal strategy expands this median estimation to include the reference sample differences at all the spatial positions, weighting each difference according to its spatial distance to \mathbf{p} . As the influence of the reference samples over \mathbf{p} decreases with their distance to \mathbf{p} , these weights are obtained through the evaluation of spatial Gaussians at these distances:

$$w_i \propto \exp \left\{ - \frac{[(r - r_i)^2 + (c - c_i)^2]}{2h_s^2} \right\},$$

$$i \in \{1, \dots, N_\beta\}. \quad (4)$$

4 Foreground Modeling

The likelihood of the foreground, φ , is estimated as a mixture of a uniform colorspace distribution and a kernel density function.² Using N_φ reference samples, \mathbf{z}_i , obtained from the detections along the last T_φ images, this likelihood is defined as:

$$\hat{f}(\mathbf{p}|\varphi) = \alpha\gamma + \frac{(1-\alpha)}{N_\varphi(2\pi)^{5/2}|\mathbf{H}_\varphi|^{1/2}}$$

$$\times \sum_{i=1}^{N_\varphi} \prod_{j=1}^5 \exp \left\{ - \frac{[\mathbf{p}(j) - \mathbf{z}_i(j)]^2}{2\mathbf{H}_\varphi(j,j)} \right\}, \quad (5)$$

where $\mathbf{H}_\varphi = \text{diag}(h_{\varphi,R}^2, h_{\varphi,G}^2, h_{\varphi,B}^2, h_{\varphi,r}^2, h_{\varphi,c}^2)$, $\alpha \in (0, 1)$ is a mixture factor, and γ is the constant density of a uniform random variable in the bounded five-dimensional set of components.

In addition, to improve the quality of the results, the spatial coordinates of the foreground reference samples are updated from each image to the next one to compensate for the displacement of the moving foreground objects.⁷

This update is done by a multi-region particle filter that is able to deal with appearing and disappearing foreground objects. As a consequence, the updated reference samples are distributed as color-spatial homogeneous foreground regions corresponding to the real modes of the foreground probability density function.

Therefore, assigning to each foreground sample a bandwidth matrix adapted to the mode it belongs to, the foreground model can be accurately estimated. To obtain these bandwidth matrices we propose to cluster these homogeneous regions by the application of a five-dimensional mean shift based segmentation.

Unlike most approaches using mean shift,⁸ to improve the quality of the resulting clustering, we use kernels that consider different bandwidth values for each dimension and benefit from the use of the previously estimated bandwidth matrices, \mathbf{H}_φ . Taking into account these considerations, foreground clusters can be obtained by iteratively applying on each reference sample the vector:

$$v(\mathbf{z}) = \left[\sum_{i=1}^{N_\varphi} \mathbf{G}(\mathbf{z} - \mathbf{z}_i) \mathbf{z}_i \right] \left[\sum_{i=1}^{N_\varphi} \mathbf{G}(\mathbf{z} - \mathbf{z}_i) \right]^{-1} - \mathbf{z}, \quad (6)$$

where \mathbf{G} is the negative derivative of the profile of the kernel. We propose to use Epanechnikov kernels, as they allow a very fast computation of $v(\mathbf{z})$. Therefore, $v(\mathbf{z})$ is finally calculated as:

$$v(\mathbf{z}) = \frac{1}{|\Psi|} \sum_{i \in \Psi} (\mathbf{z} - \mathbf{z}_i), \quad (7)$$

where $\Psi = \left\{ i \in (1, \dots, N_\varphi) \mid \left| \frac{\mathbf{z}_i - \mathbf{z}}{\mathbf{H}_\varphi^{1/2}} \right| \leq 1 \right\}$,

where $|\Psi|$ is the cardinality of the set Ψ .

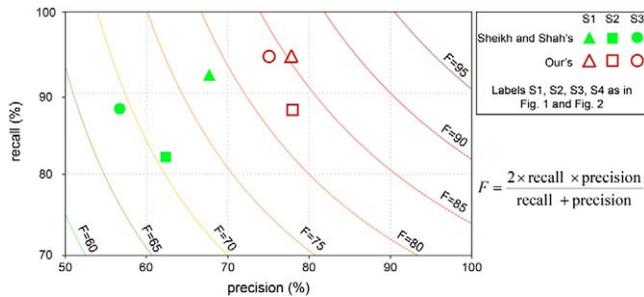


Fig. 2 Quantitative comparison between Ref. 3 and our proposal. The curve lines represent some isopercentages of the figure of merit F , which evaluates jointly the recall and the precision.

5 Results

The proposed strategy has been tested on a great variety of scenarios. Here, a selection of three sequences representing three challenging situations (e.g., dynamic backgrounds or moving objects similar to background regions) has been used to show the performance of our proposal.

Our algorithms have been compared with an outstanding background-foreground non-parametric modeling approach.³ It makes use of fixed diagonal bandwidth matrices in a spatio-temporal set of components, and could be considered the reference spatio-temporal non-parametric modeling strategy.

Figure 1 presents some partial and final results where the proposed strategy [Fig. 1(ii)] is compared with the reference method [Fig. 1(i)]. These results show that the proposed dynamical estimation strategies provide smoother background models and more accurate foreground likelihoods, resulting in improved final detections.

Finally, recall-precision percentages and F values are provided in Fig. 2. These results show that the proposed methods significantly augment the F values, because they reduce the amount of false alarms (higher precision) and increase the amount of correctly detected foreground pixels (higher recall).

6 Conclusion

An innovative dynamic kernel bandwidth estimation strategy has been presented, which is suitable for spatio-temporal

non-parametric background and foreground modeling schemes. In the background modeling, kernel bandwidth is estimated from a robust statistical analysis of spatially weighted differences between reference samples. On the other hand, kernel bandwidth for the foreground modeling is estimated through a mean shift based analysis over previously detected foreground regions that have been spatially updated using a particle filter-based tracking strategy.

Our proposal has been compared with an outstanding reference moving object detection approach. The obtained results clearly show the excellent capabilities of our approach, as it reduces the amount of false positives, whereas increasing the amount of correctly detected foreground pixels.

Acknowledgments

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

References

1. J. E. Ha and W. H. Lee, "Foreground objects detection using multiple difference images," *Opt. Eng.* **49**(4), 7201-1-5 (2010).
2. M. Tavakkoli et al., "Non-parametric statistical background modeling for efficient foreground region detection," *Mach. Vis. Appl.* **20**(6), 395-409 (2009).
3. Y. Sheikh and M. Shah, "Bayesian modeling of dynamic scenes for object detection," *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(11), 1778-1792 (2005).
4. M. Cristani et al., "Background subtraction for automated multisensor surveillance: a comprehensive review," *EURASIP J. Adv. Signal Process.* **343057**, 1-24 (2010).
5. J. G. Liao, Y. Wu, and Y. Lin, "Improving Sheather and Jones' bandwidth selector for difficult densities in kernel density estimation," *J. Nonparametr. Stat.* **22**(1), 105-114 (2010).
6. A. Elgammal et al., "Background and foreground modeling using non-parametric kernel density estimation for video surveillance," *Proc. IEEE* **90**(7), 1151-1163 (2002).
7. C. Cuevas and N. García, "Moving object detection for real-time high-quality lightweight applications on smart cameras," *IEEE Int. Conf. Consumer Electronics*, IEEE, Las Vegas (NV), USA, pp. 479-480 (2011).
8. M. Nieto et al., "Line segment detection using weighted mean shift procedures on a 2D slice sampling strategy," *Pattern Anal. Appl.* **14**(2), 149-163 (2011).