TRACKING-BASED NON-PARAMETRIC BACKGROUND-FOREGROUND CLASSIFICATION IN A CHROMATICITY-GRADIENT SPACE

Carlos Cuevas and Narciso García

Grupo de Tratamiento de Imágenes - E.T.S. Ing. Telecomunicación
Universidad Politécnica de Madrid - Madrid - Spain
{ccr, narciso}@gti.ssr.upm.es

ABSTRACT

This work presents a novel background-foreground classification technique based on adaptive non-parametric kernel estimation in a color-gradient space of components. By combining normalized color components with their gradients, shadows are efficiently suppressed from the results, while the luminance information in the moving objects is preserved. Moreover, a fast multi-region iterative tracking strategy applied over previously detected foreground regions allows to construct a robust foreground modeling, which combined with the background model increases noticeably the quality in the detections. The proposed strategy has been applied to different kind of sequences, obtaining satisfactory results in complex situations such as those given by dynamic backgrounds, illumination changes, shadows and multiple moving objects.

***Index Terms*—** Non-parametric modeling, background, foreground, segmentation, normalized components, gradients, tracking

1. INTRODUCTION

Nowadays, the increased computational speed of processors enables new applications of vision technology in several fields [1]: video-surveillance, monitoring, people motion analysis, human-machine interaction, and video coding based on objects (MPEG-4). In these applications, the detection of unusual motion is a key step for high level object analysis tasks such as object detection, tracking, classification, and event analysis.

Background subtraction techniques are commonly used in order to achieve high sensitivity in the detection of moving objects with the lowest possible number of false alarms rates. These techniques try to efficiently estimate a background model from a temporal sequence of images, and can be evaluated according to different criteria: speed (computational cost), memory requirements, and accuracy.

Some approaches aim to maximize the speed and to reduce the memory requirements, hence, they are suitable for short sequences where no important changes occur [1]. Nevertheless, these methods are not efficient in the presence of noise, illumination changes, and non-static backgrounds (containing rain, trees, flags, etc.).

To solve these limitations many multimodal methods have been developed throughout the last years. These methods are able to model more than one state for each pixel, classifying correctly as background objects the elements with cyclic movements. One key reference is the Mixture of Gaussians (MoG) model proposed by Stauffer and Grimson [2], widely used for tasks such as object detection and tracking [3]. This model makes use of a combination of Gaussians to obtain an adaptive model of each one of the pixels. Other techniques try to model the background variations by representing the changes in the scene with different states (day or night, sun or rain, etc) making use of Hidden Markov Models (HMM) [4].

While these techniques solve some complex situations with multimodal backgrounds, they have some limitations. MoG are not flexible enough to model complex density functions since they typically require the maximum number of components in the mixture in advance. Additionally, they are noticeably slow when more than 5 Gaussians are used to model each image pixel. As for HMM, it is difficult to select an appropriate model, the initialization is complex and, moreover, they include a slow training system.

More recently, non-parametric approaches methods have been proposed to tackle the background subtraction problem in environments where background statistics at the pixel level cannot be described parametrically. These do not consider the values of the pixels as a particular distribution, and build a probabilistic representation of the observations using a recent sample of values for each pixel [5] [6] [7]. However, the main drawback of these non-parametric approaches is that, for each pixel in each frame, the average of all the kernels centered at each sample should be computed, resulting in very high memory and computational requirements [6]. In addition, the size of a temporal window needs to be specified (bigger windows can improve the results, but they involve higher memory requirements and computational cost). Moreover, spatial dependencies are not exploited and the presence of shadows is usually incorrectly classified, consequently, decreasing the quality of the results.

In this work, a novel strategy for moving object detection is presented. Non-parametric background and foreground models are obtained by using an innovative combination of chromaticity and gradients. This combination reduces the impact of shadows in the classification whereas maintains the importance of the luminance information, which would be lost by only using chromaticity information. Besides, to improve the results obtained by methods based on exclusively a background modeling, here it is proposed a new alternative to model the foreground. This is based on the update of previously detected foreground regions, through a multi-region iterative tracking algorithm based on Mean-Shift. In this way, robustness against noise in enhanced and a larger number of pixels are correctly classified, improving the overall quality of the results. Additionally, high quality results are obtained in complex situations such as those with dynamic background or illumination changes.

The paper is organized as follows. Section 2 presents the non-parametric background model and describes the space of components applied to the model. In Section 3 the foreground modeling

This work has been partially supported by Alcatel-Lucent and the Spanish Administration agency CDTI under project CENIT-I3media and by the Ministerio de Ciencia e Innovación of the Spanish Government under project TEC2007-67764 (SmartVision).

and the multi-region tracking algorithm are detailed. Sections 4 and 5 show the results and the conclusions, respectively.

2. BACKGROUND MODEL

2.1. Non-parametric estimation

Let us consider that each pixel, at time t, is defined as a d-dimensional vector $\mathbf{x}_t \in \mathbb{R}^d$. The information contained in this vector is described in section 2.2. Each pixel has an associated set of neighbor pixels, $\{\mathbf{x}_i\}_{i=1}^N$, defined by a spatial bandwidth, σ_{β_s}, in the T_β previous images. The probability density function (pdf) that the pixel $\mathbf{x}_t$ belongs to the image background, β, can be non-parametrically estimated as in [7]:

$$Pr(\mathbf{x}_t|\beta) = \frac{1}{N}\sum_{i=1}^{N} K_H\left(\mathbf{x}_t - \mathbf{x}_i\right) \tag{1}$$

with the kernel estimator, K_H, defined as:

$$K_H\left(\mathbf{x}\right) = \frac{1}{\|H\|^{\frac{1}{2}}} K\left(\frac{\mathbf{x}}{H^{\frac{1}{2}}}\right) \tag{2}$$

where H is a symmetric positive definite $d \times d$ covariance matrix that specifies the "width" of the kernel around each sample point $\mathbf{x}$. Typically, a Gaussian is chosen as the kernel estimator function, K_H, due to its continuity, differentiability and locality properties [6], although other functions can be employed (e.g. triangular, uniform, Epanechnikov, etc). In this work, a d-variate Gaussian kernel function has been used. Therefore, the density can be estimated as:

$$\begin{aligned} &Pr\left(\mathbf{x}_t|\beta\right) = \\ &= \frac{1}{N}\sum_{i=1}^{N} \frac{1}{(2\pi)^{\frac{d}{2}}\,\|H\|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x}_t-\mathbf{x}_i)^\top H^{-1}(\mathbf{x}_t-\mathbf{x}_i)\right) \end{aligned} \tag{3}$$

The selected covariance matrix, H, is very important for kernel density estimation [8]. Numerous approaches have been proposed with different alternatives looking for practical and fast algorithms. In this work an alternative for adaptive background density estimation, which is an evolution of that proposed by Mittal in [5], is presented. Mittal proposes a hybrid density estimator where H is a function of the sample point, $\mathbf{x}_i$, and the estimation point, $\mathbf{x}_t$:

$$H\left(\mathbf{x}_t, \mathbf{x}_i\right) = \Sigma_{\mathbf{x}_i} + \Sigma_{\mathbf{x}_t} \tag{4}$$

where $\Sigma_{\mathbf{x}_i}$ and $\Sigma_{\mathbf{x}_t}$ are the covariance matrices of the sample point $\mathbf{x}_i$ and the estimation point $\mathbf{x}_t$, respectively. As a fully parameterized H increases the complexity of the estimation [8], the covariance matrices are chosen for this approach as diagonal matrices containing the corresponding bandwidth for each dimension:

$$\begin{aligned} \Sigma_{\mathbf{x}_i} &= diag\left(\sigma_{i,1}^2, \sigma_{i,2}^2 \ldots \sigma_{i,d}^2\right) \\ \Sigma_{\mathbf{x}_t} &= diag\left(\sigma_{t,1}^2, \sigma_{t,2}^2 \ldots \sigma_{t,d}^2\right) \end{aligned} \tag{5}$$

thus the density estimation is reduced to:

$$\begin{aligned} &Pr\left(\mathbf{x}_t|\beta\right) = \\ &= \frac{1}{N}\sum_{i=1}^{N}\prod_{j=1}^{d} \frac{1}{\sqrt{2\pi}\left(\sigma_{t,j}^2+\sigma_{i,j}^2\right)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\frac{\left(x_{t,j}-x_{i,j}\right)^2}{\sigma_{t,j}^2+\sigma_{i,j}^2}\right) \end{aligned} \tag{6}$$

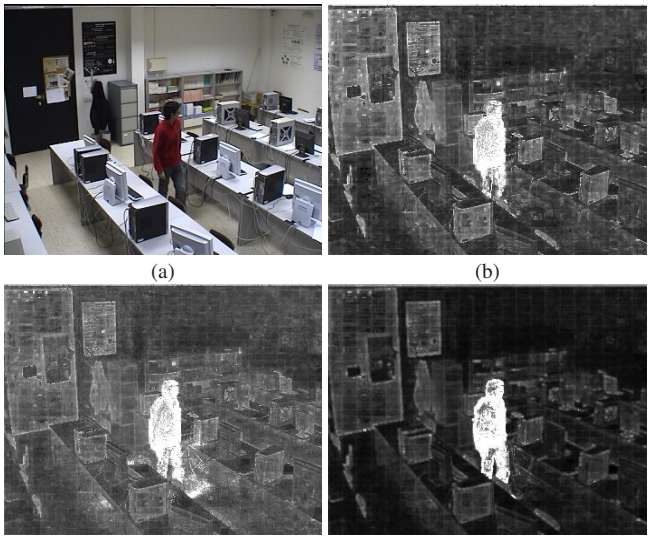


Fig. 1. Background modeling through the combination of normalized components and gradients. Whiter values correspond to lower background densities. (a) Original frame, (b) Background log-negative pdf using only chromaticity components, (c) Background log-negative pdf using chromaticity and saturation, (d) Background log-negative pdf combining normalized components and gradients.

2.2. Observations - Space of components

The presence of shadows in the scene is very common. Working in a typical RGB color space, they are frequently wrongly detected as part of the foreground, resulting in a source of confusion in posteriori phases of analysis [6]. To avoid this problem and discriminate between moving objects and their shadows, the chromaticity (normalized components) can be used:

$$R_n = \frac{R}{R+B+G} \qquad G_n = \frac{G}{R+B+G} \tag{7}$$

where R, G and B are the RGB values observed at a pixel.

Under certain conditions, the normalized components are invariant to illumination changes [5] and they allow shadow suppression from the foreground objects. Nevertheless, these components have the disadvantage of losing lightness information. To address these limitation, some authors [5] [6] add the saturation, $s = R+G+B$, to the pair of normalized components.

Although color and lightness information changes are successfully detected by using the normalized components jointly to the saturation, shadows are erroneously classified again as foreground areas. To solve this limitation and detect correctly both color and illumination changes in the proposed strategy, a novel and effective combination of chromaticity and gradients has been used to construct the non-parametric model. Combining the two normalized components, their gradients, and the gradient of s, robust and quasi-illumination invariant results are obtained.

To preserve all the gradient information, both horizontal and vertical components are maintained. Moreover, to exploit the spatial information of the samples, the concept of spatial persistence, proposed in [8], has been employed adding the coordinates of the pixels, (x, y), to the set of chromaticity and gradients. For these spatial components, a fixed bandwidth, σ_{β_s}, has been selected. Hence, the proposed feature vector is defined by $d = 10$ components:

$$\mathbf{x} = (R_n, G_n, h_{R_n}, v_{R_n}, h_{G_n}, v_{G_n}, h_s, v_s, x, y) \tag{8}$$

where h and v are respectively the horizontal and vertical gradients of chromaticity and saturation, (R_n, G_n, s).

While typical color-based systems are susceptible to shadows and rushed illumination changes, gradients are relatively less affected by changes in the lightness information [1] and can be combined with color effectively. Moreover, when background and foreground have the same color, it is difficult to discriminate them. Therefore, a natural approach to model their characteristics is to use color combined with gradient. Another obvious advantage is that gradients provide a higher level of object representation than just the color components.

Figure 1 shows an example where the background pdf has been obtained using different sets of components. An object, with color and gray areas, moves generating variable shadows. Using only the normalized components, shadows are correctly classified but the results are not satisfactory in the areas of the moving object without enough color information. However, although the results are better combining chromaticity and saturation, shadows are wrongly detected as foreground. Nevertheless, if saturation is substituted by the set of proposed gradients, shadows are finally removed. Note that, in this last case, lightness information is taken into account and the contrast is much better.

3. FOREGROUND MODEL

Sometimes moving objects show similar colors to those of the background. In these cases object detection using only background modeling is not sufficient to distinguish between foreground and background [9]. To address this limitation and to enhance detection in general, an efficient and novel foreground modeling has been developed in this work. As will be shown, the combination of this foreground modeling and the previously presented background modeling yields more accurate detections.

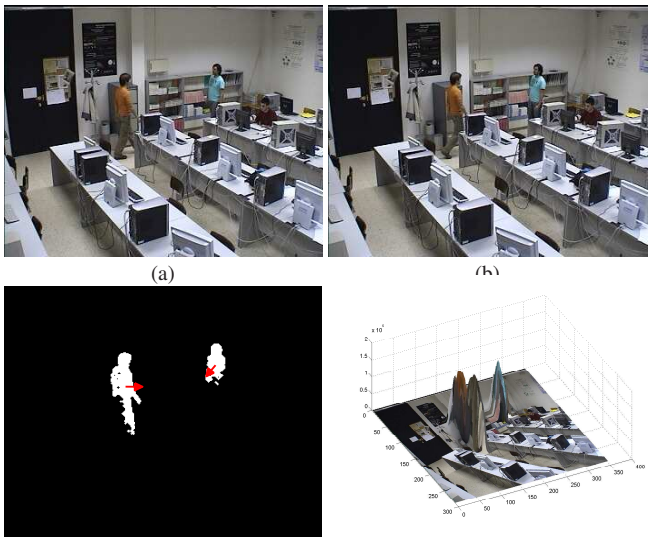
The main contribution to the foreground modeling is the proposal of a multi-region tracking strategy based on iterative approximations. At each new image, through iterative region-based comparisons [10], a search for the previously detected foreground objects is done. In this way, the coordinates of the foreground information can be updated along the sequence. This update, maximizes the influence of the foreground data over the objects they belong to and reduces their influence in other regions, resulting in an efficient and robust foreground modeling. Additionally, the estimation of the positions of the foreground objects reduces the amount of data required by the model, reducing significantly the memory requirements and the computational cost.

3.1. Non-parametric estimation

The pdf that a pixel $\mathbf{x}_t$ belongs to the foreground, Φ, can be expressed as a mixture of a uniform function and a kernel density function [8] as:

$$Pr\left(\mathbf{x}_t|\Phi\right) = \alpha\gamma + (1-\alpha)\,\frac{1}{M}\sum_{i=1}^{M} K_{H_\Phi}\left(\mathbf{x}_t - \mathbf{x}_i\right) \qquad (9)$$

where K_{H_Φ} is a multi-dimensional kernel function similar to the one defined for the background model in (6), M is the number of data, H_Φ is the foreground bandwidth matrix, $\alpha << 1$ is the mixture weight, and γ is a random variable with uniform probability in the components defined for the feature vector. A fixed bandwidth, σ_{Φ_s}, has been selected for the spatial components and a bandwidth selection, similar to that presented previously for the background, has been used for the color and gradient components.



(a) (b) (c) (d)

Fig. 2. Foreground modeling. (a) Reference frame, (b) Current frame, (c) Reference foreground regions with their associated motion vectors, (d) Current foreground probability density function.

3.2. Tracking

Let us consider a foreground region, corresponding to a moving object detected at instant t. All the pixels of this region will be used to construct the foreground model as defined in (9) along the T_Φ following images. Usually, color and gradient information of this region will maintain similar values along the next images, thus allowing to obtain adequate foreground models. Nevertheless, as the position of the object is changing with time, the selected spatial bandwidth, σ_{Φ_s}, should be high enough to cover the current positions of the moving region, increasing noticeably the amount of used information. Moreover, other regions (static or not) that contain similar information to the foreground object could be undesirably affected by the data. In order to avoid these limitations and to increase the influence of the data over the objects they belong to, in the proposed model the spatial positions of the moving objects are updated. From each image to the following one, their displacement is obtained through an iterative tracking strategy based on Mean Shift [10].

For each one of the regions to be tracked, a target model represented by a probability density function in the feature space conformed by color and gradients, $\hat{q}$, is defined. In addition, similar candidate models, $\hat{p}(x,y)$, which depend on their spatial coordinates (x,y), are defined in the current image. To find the location of the foreground regions in the current image, the distance between target and candidates models should be minimized. This distance is defined as:

$$d\left(x,y\right) = \sqrt{1-\rho\left(\hat{p}(x,y),\hat{q}\right)} \qquad (10)$$

where $\rho\left[\hat{p}(x,y),\hat{q}\right]$ is the Bhattacharyya coefficient between $\hat{q}$ and $\hat{p}(x,y)$.

The main difference between the strategy presented in [10] and that proposed here is that the target model is constructed using only foreground information. Thus, the accuracy in the motion estimation is increased, not including background information in the model.

Figure 2 illustrates the abovementioned concepts. The image at time $t-1$ is shown in (a) and the image at time t, in (b). The

foreground regions of image at $t-1$ are shown in (c) joint with their associated motion vectors computed with the described tracking strategy. These motion vectors are used to compensate the displacement between regions in these two consecutive frames in order to build the foreground model. Finally, (d) depicts a 3D visualization of the resulting foreground pdf, where, as observed, the higher values correspond to the pixels belonging to the moving objects.

4. RESULTS

The proposed moving object segmentation strategy has been tested on different kinds of video sequences containing critical aspects, such as multiple moving objects with significant gray level areas and similar to background objects, shadows, background with non-static objects (trees, water...) and illumination changes.

A buffer of $T_\beta = 120$ images and a spatial bandwidth of $\sigma_{\beta_s} = 16$ pixels have been used to model the background. In the case of the foreground model, the update of the spatial information allows to select a lower buffer and a lower spatial bandwidth, providing satisfactory results with $T_\Phi = 10$ images and $\sigma_{\beta_s} = 4$ pixels.

Figure 3 shows some results that compare the classic Mixture of Gaussians method (using 5 Gaussians per pixel in a RGB color space) with the proposed non-parametric strategy. The first column of images corresponds to an indoor sequence where the most critical aspects are the shadows and the similarity between some area in the moving object and the background. In the rest of columns, different outdoor scenarios, which contain non-static background elements, are presented. In the results obtained through the proposed algorithms, the total number of pixels correctly classified is much better, while the number of false detections is drastically reduced. These improvements are more noticeably in situations such as illumination changes or scenes with non-static background. Moreover, the combination of gradients and chromaticity results in a enhanced classification where shadows are practically removed. The foreground model addition results in an improved moving objects segmentation where the foreground regions are more compact and accurate. In the case of the last column of images in the figure, results show the better adaptation of the non-parametric strategy when an important illumination change occurs.

5. CONCLUSIONS

This paper presents a novel and efficient background-foreground classification strategy, which delivers high quality segmentations by combining non-parametric background and foreground models. The combination of chromaticity and gradients avoids shadows, which can corrupt the segmentation results, while preserving the luminance information of the moving objects. The incorporation of a fast and innovative foreground model, where the spatial information of the data is updated according to motion information, reinforces the presence of the moving regions and reduces significantly the computational requirements. This model is obtained through the application of an efficient multi-region tracking strategy over previously detected foreground regions.

The described methods have been tested on multiple kinds of video sequences containing critical aspects such as shadows, dynamic backgrounds, or illumination changes. The results show satisfactory classifications where shadows are removed from the detections, foreground regions are compact, and the number of false positives have been significantly reduced compared to the results obtained with other methods.

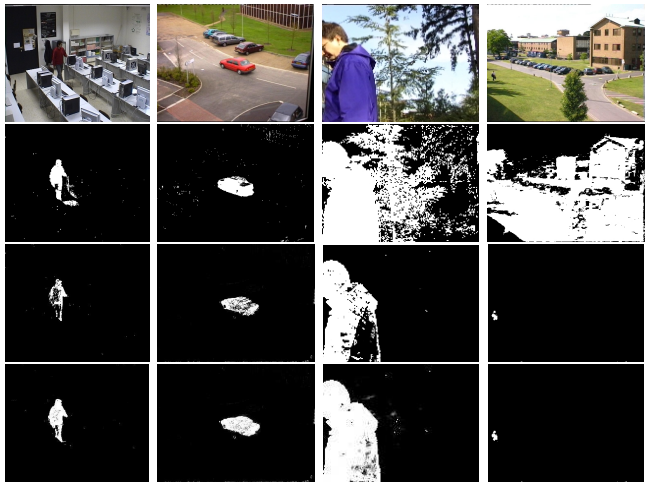


Fig. 3. First row: images from the test sequences, second row: results obtained using the MoG method, third row: results obtained by applying the proposed background model, last row: results after the addition of foreground modeling.

6. REFERENCES

[1] S.Y. Elhabian, K.M. El-Sayed, and S.H. Ahmed, "Moving object detection in spatial domain using background removal techniques - State-of-art," *Recent Patents on Computer Science*, vol. 1, no. 1, pp. 32–54, 2008.

[2] C. Stauffer and W.E.L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 747–757, 2000.

[3] M. Nieto, C. Cuevas, and L. Salgado, "Measurement-based reclustring for multiple object tracking with particle filters," in *IEEE Int. Conf. Image Processing*, 2009, pp. 4097–4100.

[4] M. Bicego, M. Cristani, and V. Murino, "Unsupervised scene analysis: A hidden Markov model approach," *Computer Vision and Image Understanding*, vol. 102, no. 1, pp. 22–41, 2006.

[5] A. Mittal and N. Paragios, "Motion-based background subtraction using adaptive kernel density estimation," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2004, vol. 2.

[6] 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.

[7] A. Tavakkoli, M. Nicolescu, G. Bebis, and M. Nicolescu, "Non-parametric statistical background modeling for efficient foreground region detection," *Machine Vision and Applications*, vol. 20, no. 6, pp. 395–409, 2009.

[8] 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.

[9] X. Zhang and J. Yang, "Foreground segmentation based on selective foreground model," *Electronics Letters*, vol. 44, pp. 851, 2008.

[10] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 564–577, 2003.