HIERARCHICAL AND SPATIAL-COLORIMETRIC MODEL TO DETECT MOVING TARGETS

C. Gabard[1], C. Achard[2], L. Lucat[1] and P. Sayd[1]
[1] *CEA, LIST, Vision and Content Engineering Laboratory, Point Courrier 94, F-91191 Gif-sur-Yvette, France*
[2] *UPMC Univ Paris 06, Institut des Systmes Intelligents et de Robotique (ISIR), Paris, France*

Keywords: MOG, SMOG, SGMM, Background Subtraction, Tracking, Foreground and Object Detection.

Abstract: Background subtraction is often one of the first tasks involved in video surveillance applications. Classical methods only use temporal modelling of the background pixels. Using pixel blocks with fixed size allows robust detection but these approaches lead to a loss of precision. We propose in this paper a model of the scene which combines a temporal and local model with a spatial model. This whole representation of the scene both models fixed elements (background) and mobile ones. This allows improving detection accuracy by transforming the detection problem in a two classes classification problem.

1 INTRODUCTION

With the explosion of the video surveillance deployment, ever more complex systems are needed to automatically detect and interpret events. These high-level processes are generally based on preliminary motion detection and tracking steps. Conventional moving area detection methods, such as (Stauffer and Grimson, 1999), exploit a statistical modeling of background pixels. Thus, spatial consideration aiming to take into account the object compactness is introduced during the post-processing step. Moreover, most of the literature approaches only use a background model. The proposed method is based on a hierarchical model of the scene. It combines the advantages of temporal pixel modeling and a global modeling of the scene which takes into account the spatial pixel consistency. The detection decision, which is no longer local but is performed on consistent pixel sets, is thereby more robust.

2 RELATED WORK

The *SGMM* approach is mainly used in two ways:

Tracking: the *SGMM* models only moving targets. Each target is then represented by a set of modes that can be temporally tracked (Wang et al., 2007; Gallego et al., 2009).

Background Subtraction: SGMM is used to model the whole scene, both background and foreground (Dickinson et al., 2009; Yu et al., 2007).

In this approach, the mixture of Gaussian is used to model the whole scene and not each pixel independently. Thus, a list of modes is update to represent the whole image. Observation at each pixel is composed of both spatial and color components $\mathbf{x}_t = [x, y, R, G, B]$. The scene is then represented by a set of Gaussian distributions in a five-dimensional space. Initialization of *SGMM* which does not require any learning period, is performed on the first frame of the sequence. For example some authors (Yu et al., 2007; Wang et al., 2007) use the Exception Maximisation algorithm. As such algorithms may slowly converge, some other works like (Dickinson et al., 2009) involve some heuristics to successively cut and merge modes by analyzing the statistical characteristics of data distribution. Finally, while an initial hand-made object segmentation is often provided (Yu et al., 2007; Wang et al., 2007; Gallego et al., 2009), some works such as (Dickinson et al., 2009) allow to dynamically create object modes.

3 THE PROPOSED APPROACH

The proposed system takes some advantages of the work presented in (Dickinson et al., 2009) and intro-

Gabard C., Achard C., Lucat L. and Sayd P..
HIERARCHICAL AND SPATIAL-COLORIMETRIC MODEL TO DETECT MOVING TARGETS.
DOI: 10.5220/0003871105050508
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2012), pages 505-508
ISBN: 978-989-8565-03-7

duces major extensions. First, we propose to monitor the mode evolutions in order to avoid important drift. A new decision step is also introduced to label each mode as background or object.

3.1 Initial Model Construction

A mode list is built from the first image to model the scene. It is necessary to check that each mode is consistent, both in terms of color and spatial compactness. In order to build the initial model, only one mode is first created using the whole image data. The settings of the component j are estimated from the pixels $x \in L_j$ according to:

$$\omega_j = \frac{n_j}{N} \qquad \mu_j = \frac{1}{n_j}\sum_{x \in L_j} x$$
$$\Sigma_j = \frac{\sum_{x \in L_j} x^T x}{n_j} - \mu_j^T \mu_j \tag{1}$$

where n_j is the number of pixels in L_j, and N the total number of pixels. A succession of split and merge operations are then performed.

3.2 Mode Splitting and Merging

Modes are split iteratively until two alternating stop criterions are reached. At each iteration, the mode maximizing the curent tested criterion is selected and if is above a predefined threshold, then the component is cut in the largest eigenvector direction. The two criterions measures respectively the color dispersion and the inverse spatial density:

$$C_j^C = \sqrt{\sum_{k=1}^{3}\left(\lambda_{j,k}^C\right)^2} \qquad C_j^S = \frac{\sqrt{\sum_{k=1}^{2}\left(\lambda_{j,k}^S\right)^2}}{n_j} \tag{2}$$

with $\lambda_{j,k}^{C|S}$ are the eigenvalues estimated from the color and spatial covariance matrix.

Afterwards similar components are gathered as described in (Dickinson et al., 2009). All pairs of components are considered and the pairs are grouped if their characteristics are similar: the average value of one mode is well represented within a confidence interval by the other mode.

Finally, components representing spatially disconnected regions are identified and split to represent these regions independently.

3.3 Pixel Assignment and Update

Once the scene model is initialized, continuous frame-based processing is performed. Given a new incoming image, the value of a pixel can be classified according to the mode (or component) providing the maximum posterior probability, C_{map}. Using the log-likelihood:

$$C_{map} = argmax_j\{log(p(\mathbf{x}_t|\theta_{(j,t)}))\} \tag{3}$$

where $p(\mathbf{x}_t|\theta_{(j,t)})$ is the probability density of the mode j with parameters $\theta_{(j,t)}$. The model is simplified by assuming color and spatial independence.

The pixel is assigned to the mode providing the highest likelihood. An uniform distribution is added to manage emergence of a new object. Regions containing an high density of pixels assigned to this distribution are used to create new components.

After pixel assignment, parameters of the components are re-estimated. For each component j, a set of parameters $\theta_{(j,sm)}$ is calculated from the current image pixels assigned to it. From the above parameters $\theta_{(j,t-1)}$, the new values $\theta_{(j,t)}$ are calculated using an adaptive learning (α_j denotes the learning rate):

$$\theta_{(j,t)} = \alpha_j.\theta_{(j,sm)} + (1-\alpha_j)\theta_{(j,t-1)} \tag{4}$$

Without constraint, modes are likely to highly deviate away from the initial model. In order to avoid mode drifts, we propose to adjust the modes by performing the same split and merge steps as during initialization. These few operations help to maintain a consistent list of components for both color and space point of view, and greatly improve detection results.

3.4 Mode Labeling

Each mode can represent a background area or be attached to a moving object. Detection output involves a *background* or *foreground* classification decision. We propose in this paper to perform the decision at the mode level using a pixel-based classification provided by a local temporal modeling like the Stauffer and Grimson (Stauffer and Grimson, 1999) algorithm. The decision is globally performed on a pixel group. To this end, we use the probability map generated by the Stauffer and Grimson method and threshold the average value of this map over all mode's pixels.

4 RESULTS

4.1 Acquisition Noise Sensitivity

This experiment aims to asses the sensitivity of our approach to the image quality. A sequence, named ”Blue Room”, is an indoor scenario whose environment is under control. Gaussian noise, spatially and

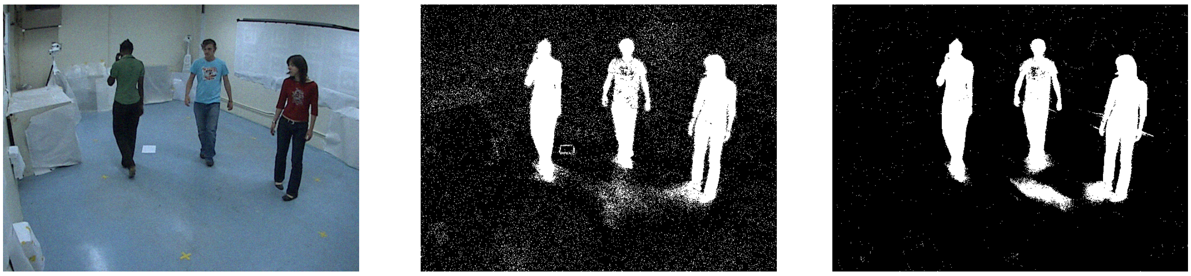

Figure 1: **"Blue Room" sequence:** Sequence is subject to a Gaussian noise with standard deviation of 10. From left to right: Source frame; Stauffer and Grimson algorithm; our algorithm.

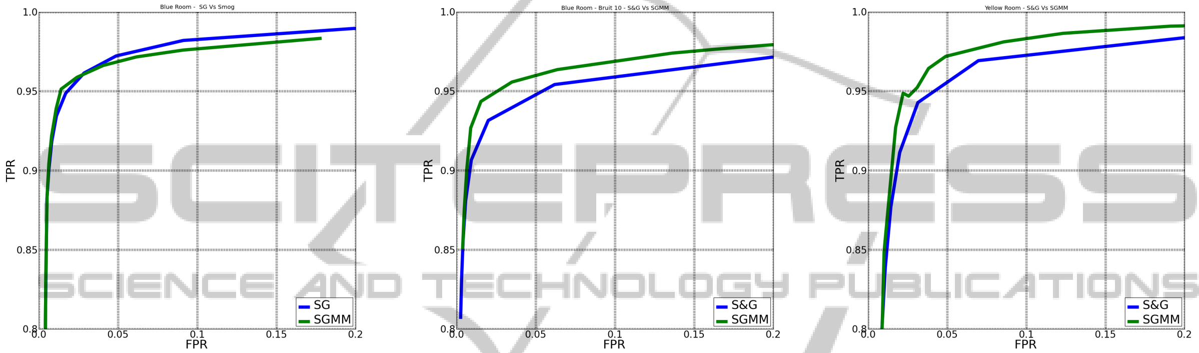

Figure 2: **ROC curves sequence results:** On left to right: "Blue Room" original sequence results; "Blue Room" degraded by a Gaussian noise with standard deviation of 10; "Yellow Room" sequence.

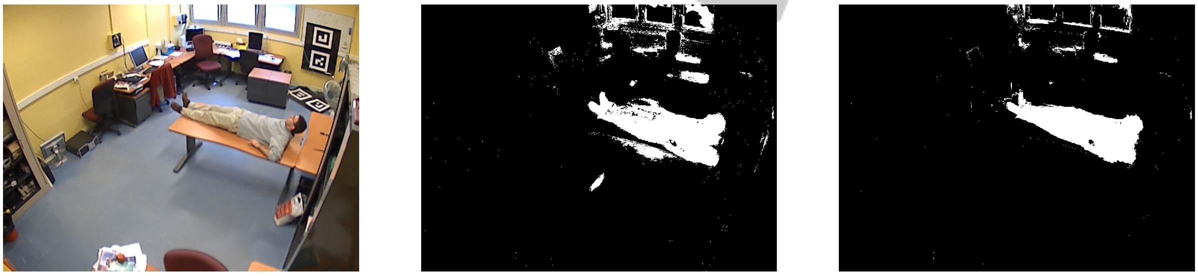

Figure 3: **"Yellow Room" sequence:** From left to right: Source frame, Stauffer and Grimson algorithm and our algorithm.

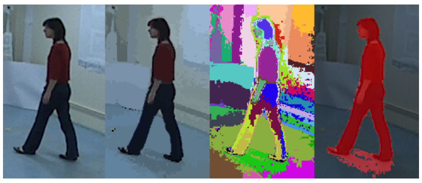

Figure 4: **Target decomposition:** Image source, associated mode mean color, mode identification and detection result.

timely independent with a standard deviation of 10, has been injected into the sequence. Some results are presented in Figure 1 and ROC curves are ploted in Figure 2. In the absence of noise, results are quite similar for both methods. However it can be observed that our algorithm performance remains stable while the Stauffer & Grimson algorithm loses precision as the noise increases.

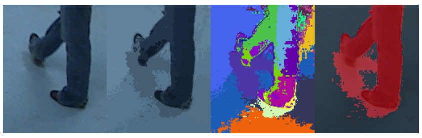

Figure 5: **Shadow separation:** Image source, associated mode mean color, mode identification and detection result.

4.2 Difficult Conditions

A more complicated sequences have been tested. "Yellow Room" is an Indoor sequence. The condi-

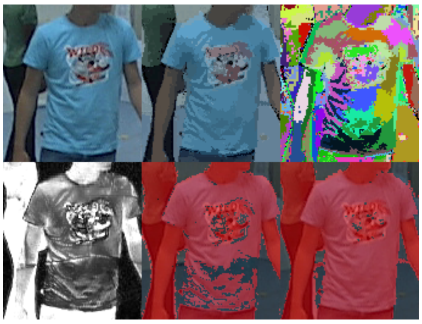

Figure 6: **Grouping robustness:** Top Line: Image source, mode mean color, mode identification; Bottom line: Probability map from ”Stauffer & Grimson”, corresponding detection results and detection result with our method.

tions are difficult due to direct outdoor lighting changing during the sequence and a light flicker. The results are plotted in Figure 3 and Figure 2. The proposed method outperforms the pixel-based method.

4.3 Detailed Analysis of some Characteristics

Beyond these general performance, it is relevant to focus on some algorithm behaviors:

Target Decomposition. Figure 4 highlights the generated modes decomposition on a person. Each mode is represented by a random color for better visualisation. It can be observed that the target and the background are clearly segmented and also that the different body members are segmented, which could be useful for a higher-level analysis.

Shadow Separation. As many algorithms, the proposed method classifies some shadows as moving objects. However, by analyzing the composition of the modes (Figure 5), we can notice that shadows are segmented in dedicated modes. Treatments for shadows removal are simplified as group of pixels can directly be studied and compared to old background.

Grouping Robustness. To illustrate the grouping interest, it is relevant to focus on difficult cases such as the Tee-shirt whose color is close to the background color (Figure 6). The Stauffer and Grimson method generates many misdetections. The mode-based decision of the proposed algorithm alleviates these problems.

Object Tracking. Finally, assuming that the movement of objects from one image to the following is relatively small, the corresponding modes follow the object. If the object movement is more important, the ”old” component will disappear and a corresponding new component will be automatically created: the segmentation and labeling process remain effective.

5 CONCLUSIONS

Conventional approaches typically performes a local modeling independently on each pixel. Using blocks of pixels of pre-determined shape and size to estimate a descriptor often leads to more robustness, but the spatial precision is then deteriorated. The proposed approach provides a solution to overcome these limitations by combining the accuracy of the pixel information with the robustness of a decision made on a set of coherent pixels. The experimental results presented emphasized that the proposed approach improves performances, with respect to both Stauffer and Grimson method and state-of-art SGMM. In particular, thanks to the spatial consistency, this approach remains very stable in noisy and dynamic conditions.

REFERENCES

Dickinson, P., Hunter, A., and Appiah, K. (2009). A spatially distributed model for foreground segmentation. *Image and vision computing*.

Gallego, J., Pardas, M., and Haro, G. (2009). Bayesian foreground segmentation and tracking using pixel-wise background model and region based foreground model. In *Image Processing (ICIP)*.

Stauffer, C. and Grimson, W. (1999). Adaptive background mixture models for real-time tracking. In *CVPR*.

Wang, H., Suter, D., Schindler, K., and Shen, C. (2007). Adaptive object tracking based on an effective appearance filter. *IEEE transactions on pattern analysis and machine intelligence*.

Yu, T., Zhang, C., Cohen, M., Rui, Y., and Wu, Y. (2007). Monocular video foreground/background segmentation by tracking spatial-color gaussian mixture models. In *Motion and Video Computing*.