A Background Modelling Algorithm for Motion Detection

Paolo Spagnolo, Marco Leo, Tiziana D Orazio, Nicola Mosca and Massimiliano Nitti

Abstract **Detecting moving objects is very important in many application contexts such as people detection, visual surveillance, and so on. The first and fundamental step of all motion detection algorithms is the background modeling. The goal of the methodology here proposed is to create a background model substantially independent from each hypothesis about the training phase, as the presence of moving persons, moving background objects, and changing (sudden or gradual) light conditions. We propose an unsupervised approach that combines the results of temporal analysis of pixel intensity with a sliding window procedure to preserve the model from the presence of foreground moving objects during the building phase. Moreover, a multilayered approach has been implemented to handle small movements in background objects. The algorithm has been tested in many different contexts, such as a soccer stadium, a parking area, a street, a beach. Finally, it has been tested even on the CAVIAR 2005 dataset.**

I. INTRODUCTION

MANY computer vision tasks require robust segmentation of foreground objects from dynamic scenes. The most used algorithms for moving objects detection are based on background subtraction. In these applications, the first and crucial step is the background creation.

Many algorithms proposed in literature in the last years present some common characteristics. Usually, independently from the applicative context, the main features that each background modeling algorithm has to handle are:

- Presence of foreground and/or moving background objects during the model building phase;
- Gradual and/or sudden variations in illumination conditions.

A first group of modeling algorithms uses statistical approaches to model background pixels. In [1,2] a pixel-wise gaussian distribution was assumed to model the background; the presence of foreground objects during the building phase could cause the creation of an unreliable model, such as in presence of light movements in the background objects, or sudden light changes. These observations suggest that probably the proposed algorithms work well in presence of a supervised training, during which ideal conditions are granted by the human interaction.

The natural evolution of these approaches was proposed in [4]: authors use a generalized mixture of gaussians to model complex non-static background. In this case the presence of foreground objects during this phase could alter the reliability of the model immediately after the creation phase.

The approach proposed in [5] was conceptually similar to that proposed in [1]. But in this work the authors did not construct a real gaussian distribution, while they preferred to maintain general statistics for each point. In this way they cope with the movements in background objects, even if they waive a correct segmentation of foreground objects in those regions. However they could encounter misdetection in presence of foreground objects during the modeling phase, and in presence of sudden light changes. The natural improvement of this approach was proposed in [6]: the basic idea of [5] was iterated in order to build a codebook for each point, providing a set of different possible values for each point. All the approaches above examined use statistical information, at different complexity level, for the background modeling.

A different category is composed by the approaches that use filters for temporal analysis. In [8] authors used a Kalman-filter approach for modeling the state dynamics for a given pixel. In [9] a non-parametric technique was developed for estimating background probabilities at each pixel from many recent samples over time using Kernel density estimation. In [10] an autoregressive model was proposed to capture the properties of dynamic scenes. An improvement of this algorithm was implemented in [11,12] to address the modelling of dynamic backgrounds and perform foreground detection. In [13] a modified version of the Kalman filter, the Weiner filter, was used directly on the data. The common assumption of these techniques was that the observation time series were independent at each pixel.

All the approaches above presented were tested on real sequences, producing interesting results, even if each of them suffered in almost one of the critical situations listed above. Moreover, most of them implicitly require a supervised background model construction.

In this work we present a background modeling algorithm able to face all the crucial situations typical of a motion detection system with an unsupervised approach; no assumptions about the presence/absence of foreground objects and changes in light conditions are required. The main idea is to exploit the pixels energy information in order to distinguish static points from moving ones. To make the system more reliable and robust, this procedure has been integrated in a sliding windows approach, that is incrementally maintained during the training phase; in this way the presence of sudden light changes and foreground objects is correctly handled, and it does not alter the final background model. In order to cope with the presence of moving background objects, a multilayered modeling approach has been implemented, combining temporal and energetic information.

The whole background creation algorithm will be explained in the next sections.

II. Background model

The main goal of a modeling algorithm is to create a reliable model limiting the memory requirements: in an ideal case the best background model could be created by observing a-posteriori all the frames of the training phase; however this solution is not reasonable then one of the constraint of our approach is to work in an incrementally mode, to reduce hardware requirements, without losing the reliability.

The implemented background modeling algorithm is based on two distinct phases, each of them tries to solve a particular modeling problem (see par. 1).

Firstly, the energy information of each image point, evaluated in a small sliding temporal window, is used to distinguish static points from moving ones. In this way we are able to obtain a statistical background model with only the contribution of background points, without the effects of foreground objects. However, with this proposed technique, the small movements of the background objects are not included in the model.

So, in order to cope with this problem, a multilayered approach has been implemented, integrating the one-layer information given by the previous step with other data deriving from a long term temporal analysis. This two operations will be explained in details in the following sections.

III. Energy information

One of the main problems of background modeling algorithm is their sensitiveness to the presence of moving foreground objects in the scene.

The proposed algorithm exploits the temporal analysis of the energy of each point, evaluated by means of sliding temporal windows. The basic idea is to analyze in a small temporal window the energy information for each point: the statistical values relative to slow energy points are used for the background model, while they are discarded for high energy points. In the current temporal window, a point with a small amount of energy is considered as a static point, that is a point whose intensity value is substantially unchanged in the entire window; otherwise it corresponds to a non static point, in particular it could be:

- a *foreground point* belonging to a foreground object present in the scene;
- a *background point* corresponding to a moving background object.

At this level, these two different cases will be treated similarly, while in the next section a more complex multilayer approach will be introduced in order to correctly distinguish between them.

A coarse-to-fine approach for the background modeling, is applied in a sliding window of size W (number of frames). The first image of each window is the coarse background model $B_c(x,y)$. In order to have an algorithm able to create at runtime the required model, instead of building the model at the end of a training period, as proposed in [3], the mean (1) and standard deviation (2) are evaluated at each frame; then, the energy content of each point is evaluated over the whole sliding window, to distinguish real background points from the other ones. Formally, for each frame the algorithm evaluates mean and standard deviation, as proposed in [2]:

$$\overline{\mu^t(x,y)} = \alpha\mu^t(x,y) + (1-\alpha)\overline{\mu^{t-1}} \qquad (1)$$

$$\overline{\sigma^t(x,y)} = \alpha\,|\,\mu^t(x,y) - \overline{\mu^t(x,y)}\,| + (1-\alpha)\overline{\sigma^{t-1}} \qquad (2)$$

only if the intensity value of that point is substantially unchanged with respect to the coarse background model:

$$\left|I^t(x,y) - B_C(x,y)\right| < th \qquad (3)$$

where *th* is a threshold experimentally selected and $I^t(x,y)$ is the intensity value of point *(x,y)* at time t.

In this way, at the end of the analysis, in the first W frames, for each point the algorithm evaluates the energy content as follows:

$$E(x,y) = \int_{t\in W} \left|I^t(x,y) - B_C(x,y)\right|^2 \qquad (4)$$

The first fine model of the background B_F is generated as:

$$B_F(x,y) = \begin{cases} (\mu(x,y), \sigma(x,y)) & if\ E(x,y) < th(W) \\ \phi & if\ E(x,y) > th(W) \end{cases} \qquad (5)$$

A low energy content means that the considered point is a static one and the corresponding statistics are included in the background model, whereas high energy points, corresponding to foreground or moving background objects cannot contribute to the model. The whole procedure is iterated on another sequence of W frames, starting from the frame *W+1*. The coarse model of the background is now the frame *W+1*, and the new statistical values (1) and (2) are evaluated for each point, like as the new energy content (4).

The relevant difference with (5) is that now the new statistical parameters are averaged with the previous values, if they are present; otherwise, they become the new statistical background model values. Formally, the new formulation of (5) become:

$$B_F(x,y) = \begin{cases} (\mu(x,y),\sigma(x,y)) \;\; if\; E(x,y) < th(W) \\ \wedge B_F(x,y) = \phi \\ \beta * B_F(x,y) + (1-\beta) * (\mu(x,y),\sigma(x,y)) \\ if\; E(x,y) < th(W) \wedge B_F(x,y) \neq \phi \\ \phi \qquad if\;\; E(x,y) > th(W) \end{cases} \tag{6}$$

The parameter β is the classic updating parameter introduced in several works on background subtraction ([1], [2], [5]). It allows to update the existent background model values to the new light conditions in the scene.

The whole procedure is iterated N times, where N could be a predefined value experimentally selected to ensure the complete coverage of all pixels. Otherwise, to make the system less dependent from any a-priori assumption, a dynamic termination criteria is introduced and easily verified; the modeling procedure stops when a great number of background points have meaningful values:

$$\#(B_F(x,y) = \phi) \cong 0 \tag{7}$$

IV. MULTILAYER ANALYSIS

The approach described above allows the creation of a statistical model for each point of the image, even if covered by moving objects. However, it is not able to distinguish movements of the background objects (for example, a tree blowing in the wind) from foreground objects. So, the resulting model is very sensitive to the presence of small movements in the background objects.

The starting point of the proposed approach is the observation that, if a foreground object appears in the scene, the variation in the pixel intensity levels is unpredictable, without any logic and/or temporal meaning. Otherwise, in presence of a moving background object, there will be many variations of approximately the same magnitude, even if these variations will not have a fixed period (this automatically excludes the possibility to use frequency-based approaches, i.e. Fourier analysis).

So, the goal of this step is to use a multilayer approach for the modelling, with the aim of discarding layers that correspond to variation exhibited only a few times for a given point. Differently, layers that in the observation period return more times will be taken (they probably correspond to static points covered by background moving objects). Formally, the main idea proposed in the previous section remain unchanged, but it is now applied to all the background layers. The concept of mean and standard deviation proposed in (1) and (2) become:

$$\overline{\mu_i^t(x,y)} = \alpha\mu_i^t(x,y) + (1-\alpha)\overline{\mu_i^{t-1}} \tag{8}$$

$$\overline{\sigma_i^t(x,y)} = \alpha\,|\,\mu_i^t(x,y) - \overline{\mu_i^t(x,y)}\,| + (1-\alpha)\overline{\sigma_i^{t-1}} \tag{9}$$

where i changes in the range [1 K], and K is the total number of layers. Similarly, for each frame of the examined sequence, the decision rule proposed in (3) for the updating of the parameters becomes:

$$\left|I^t(x,y) - B_C^i(x,y)\right| < th \tag{10}$$

where the notation i indicates the examined layer. It should be noted that, initially, there is only one layer for each point, the coarse background model (that correspond to the first frame). Starting from frame #2, if the condition (10) is not verified, a new layer is created. In this way, at the end of the observation period, for each point the algorithm builds a statistical model given by a serious of couple (μ, σ) for each layer. The criteria for selecting or discarding these values is based again on the evaluation of the energy content, but now (4) is evaluated for each layer i:

$$E^i(x,y) = \int_{t\in W} \left|I^t(x,y) - B_C^i(x,y)\right|^2 \tag{11}$$

Different layers are created only for those values that occur a certain number of times in the observation period. However, in this way both foreground objects and moving background ones contribute to the layer creation. In order to distinguish these two different cases, and maintain only information about moving background objects, the *overall occurrence* is evaluated for each layer:

$$O^i(x,y) = \#W\left|(x,y)\right. \;\; that\;\; contributes\; to\; the \;\; statistics\;\; of\; the\, layer\, i \tag{12}$$

$O^i(x,y)$ counts the number of sliding windows that contributes to the creation of the statistic values for the layer i. At this point, the first K layers with the highest overall occurrences belong to the background model, while the others are discarded.

After the examination of all the points with (12), the background model contains only information about the static background and moving background objects, while layers corresponding to foreground objects are discarded since they occur only in a small number of sliding windows.

The use of sliding windows allows to greatly reduce the memory requirements; the trade-off between goodness and hardware requirements seems to be very interesting with respect to the others proposed in [11] and [3].

V. EXPERIMENTAL RESULTS

We have tested the proposed algorithm on different sequences (archeological site, laboratory, museum, soccer stadium, beach) acquired by ourselves in different conditions, in both indoor and outdoor environments, at different frame rate. Two sequences present in the CAVIAR dataset (see CAVIAR home page) have also been considered. In our experiments, we have chosen to use a sliding window containing 100 frames, independently from the camera frame rate. The first tests were carried out to evaluate the number of layers necessary for a given situation. In table 1 the mean number of layers for each context is proved. As it can be seen, this value is smaller for more structured contexts (laboratory, soccer stadium),while it is higher in generic outdoor contexts (archeological site, CAVIAR seq.1). The maximum number of layers in our experiments was fixed to 5. The presence of moving background objects in the beach and archeological site contexts increases the number of layers. In more controlled environments, like the laboratory, requiring a small number of layers, probably the multilayer approach can be considered unnecessary.

Test Sequence	Mean number of layers
Archeological site	3.12
Laboratory	1.23
Museum	2.05
Soccer Stadium	1.92
Beach	4.33
CAVIAR seq. 1	2.28
CAVIAR seq. 2	1.54

Table 1. the mean number of layers for each of the examined different contexts

In order to have a quantitative representation of the reliability of the background models, we have chosen to test them by using a standard, consolidated motion detection algorithm, proposed in [2]. A point will be considered as a foreground point if it differs from the mean value more than two times the standard deviation:

$$\left|I(x,y) - B^{i}(x,y)\right| > 2 * V^{i}(x,y) \qquad (13)$$

Starting from the detection rule (13), we have chosen to use the perturbation detection rate (PDR) analysis to validate our approach. This technique, as explained in [7], makes the experimental results less sensitive to the effects of a manual ground truth segmentation. The goal of the PDR analysis is to measure the detection sensitivity of a background subtraction algorithm. This analysis is performed by shifting or perturbing the entire background distributions by values with fixed magnitude , and computing an average detection rate as a function of Δ. More details about this procedure can be found in relative paper. The PDR analysis has been applied to all the experimental contexts presented above. The test set is given by 500 points for each frame, 200 frames for each sequence in each context. So, for each Δ, 200*500 perturbations and detection tests were performed. In figure 1 we have plotted the resulting PDR graphs. The worst results have been obtained in the beach, where the critical conditions due to the presence of moving background objects decrease the performance. In this case, the pixel intensity variations, due to the movement of the vegetation, are amplified by the perturbation introduced, causing a decrease of the global detection ability. On the other hand, the results obtained in the remaining contexts are very interesting, with a fast growth of the curve towards best performances.

We have preferred to propose our experimental results instead of compare them with the same obtained by others because of we consider that our implementation of algorithms of other authors can be not perfect, so the obtained results could be corrupted by this incorrect implementation.

As a future work, we are including the background modeling algorithm in a complete motion detection system, able to take advantage of the main characteristics of the proposed algorithm.

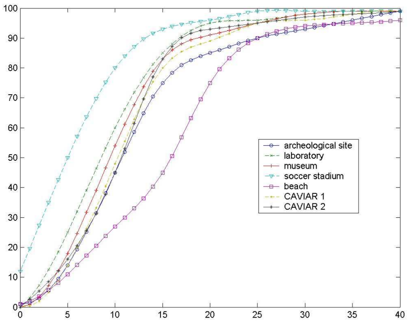

Figure 1: the PDR analysis on the test sequences. It can be note that the best performance have been obtained in the soccer stadium and in the indoor contexts, while the worst results have been reported in the archeological site, probably due to the presence of moving vegetations.

REFERENCES

[1] C.R. Wren, A. Azarbayejani, T. Darrell, A.P. Pentland, Pfinder: real-time tracking of human body , *IEEE Trans. Patt. Anal. Mach. Intell.*, 19(7), pp. 780 785, July 1997.

[2] T.Kanade, T.Collins, A.Lipton, Advances in Cooperative Multi-Sensor Video Surveillance , *Darpa Image Und. Work., Morgan Kaufmann*,pp.3-24, Nov.1998.

[3] A.J. Lipton, N. Haering, ComMode: an algorithm for video background modeling and object segmentation , Proc. of ICARCV, pages 1603- 1608, vol.3, 2002

[4] C. Stauffer, and W. Grimson, Adaptive background mixture models for real-time tracking , Proceedings of Computer Vision and Pattern Recognition, pages II 246-252, 1999

[5] I. Haritaoglu, D. Harwood, L.S. Davis, Ghost: A human body part labeling system using silhouettes , *Fourteenth Int. Conf. on Patt. Rec.*, Brisbane, Aug. 1998.

[6] K Kim, T.H. Chalidabhongse, D. Harwood, L. Davis, Background modeling and subtraction by codebook construction , Proc. of ICIP 2004, Volume 5, Pages 3061 64

[7] T.H.Chalidabhongse, K.Kim, D.Harwood, and L.S.Davis, A Perturbation Method for Evaluating Background Subtraction Algorithms ,*Proc. VS-PETS 2003,*Nice,France,Oct.11-12,2003

[8] D. Koller, J. Weber, J. Malik, Robust multiple car tracking with occlusion reasoning , in ECCV 1994, pages 189-196, Stockholm, Sweden, May 2004

[9] A. Elgammal, D. Harwood, L.S. Davis, Non-parametric model for background subtraction European Conference Computer Vision, Vol. 2, pp. 751-767, 2000

[10] G. Doretto, A. Chiuso, Y.N. Wu, and S. Soatto, Dynamic textures , International Journal on Computer Vision, 51 (2), pp 91-109, Febr. 2003

[11] A. Monnet, A. Mittal, N. Paragios, and V. Ramesh, Background modelling and subtraction of dynamic scenes , in ICCV, pp. 1305-1312, Nice, France, October 2003

[12] J. Zhong, and S. Sclaroff, Segmenting foreground objects from a dynamic, textured background via a robust kalman filter in ICCV, pp. 44-50, Nice, France, October 2003

[13] K.Toyama, J.Krumm, B.Brumitt, and B. Meyers, Wallflower: Principles and practice of background maintenance , in ICCV, pp. 255-261, Greece, Sept. 1999