ROBUST AND EFFICIENT BACKGROUND SUBTRACTION BY QUADRATIC POLYNOMIAL FITTING

Alessandro Lanza *Federico Tombari* *Luigi Di Stefano*

DEIS - ARCES, University of Bologna, Bologna, Italy

ABSTRACT

We present a background subtraction algorithm aimed at efficiency and robustness to common sources of disturbance such as illumination changes, camera gain and exposure variations, noise. The approach relies on modeling the local effect of disturbance factors on a neighborhood of pixel intensities as a second-degree polynomial transformation plus additive gaussian noise. This allows for classifying pixels as changed or unchanged by a simple least-squares polynomial fitting procedure. Experimental results prove that the approach is state-of-the-art in challenging sequences characterized by sources of disturbance yielding sudden and strong background appearance changes.

1. INTRODUCTION AND PREVIOUS WORK

Background subtraction is the first critical step in many video analysis applications, such as e.g. intelligent video surveillance and automated traffic monitoring. The main difficulty with background subtraction consists in discerning significant changes in presence of the sources of disturbance that may modify the appearance of the reference scene, such as noise, gradual or sudden illumination changes (e.g. due to time of the day or a light switch), dynamic adjustments of camera parameters (e.g. auto-exposure, auto-gain). To carry out background subtraction robustly is a challenging research issue, as vouched by the richness of proposals found in literature (see [1] for a recent survey).

A first class of algorithms relies on learning off-line, i.e. at initialization time, a per-pixel model of the background appearance and on continuously updating it to incorporate the possible effects of disturbs. Then, a pixel from a new frame is marked as foreground if its intensity "does not fit" the current background appearance model. Several statistical models, ranging from a single Gaussian [2] to mixtures of Gaussians [3] and kernel-based non-parametric models [4] have been proposed. Yet, the need for on-line updating the model implies a tradeoff associated with the chosen adaptation rate: prompt model updating can result in foreground elements being mistakenly integrated into the background model whilst slow updating makes this first class of algorithms inadequate to cope with fast background appearance changes, such as those due to sudden illumination changes.

A second class of algorithms relies on estimating off-line the background model and a-priori modeling the background appearance changes yieldable by the sources of disturbance within small image patches. Accordingly, a pixel from a new frame is classified as foreground if the intensity variations with respect to the background observed within a surrounding neighborhood "does not fit" the a-priori model of changes. Since changes due to disturbs are a-priori modeled, these algorithms can handle seamlessly slow and sudden photometric variations and the background model does not need to be updated. A critical issue with these algorithms concerns the a-priori model of changes, which is generally linear [5, 6] or order-preserving [7, 8]. In principle, the more restrictive such a model the higher is the ability to detect foreground elements (sensitivity) but the lower is robustness to sources of disturbance (specificity). As discussed in [7], many non-linearities may arise in the image formation process, so that a more liberal model than the linear is often required to achieve adequate robustness in practical applications.

In this paper we propose a novel background subtraction algorithm belonging to the second class which assumes a second-degree polynomial transformation as a-priori model for the local intensity variations due to disturbs. That is, we assume a model that is more general than the linear and less general than the order-preserving. In this manner, our approach holds the potential for achieving a better tradeoff between robustness and sensitivity.

2. PROPOSED ALGORITHM

Let B be the background and F the current frame. For each pixel in B and its correspondent in F, we take into consideration a surrounding neighborhood of N pixels that we denote as, respectively, $X = \{x_1, \cdots, x_N\}$ and $Y = \{y_1, \cdots, y_N\}$. We aim at detecting scene changes occurring in the central pixel by evaluating the local intensity information contained in X and Y.

We assume that the background model is computed by means of a statistical estimation over an initialization sequence (e.g. temporal averaging of tens of frames) so that noise affecting the inferred background intensities can be neglected. Hence, X can be thought as a vector of noiseless intensities that are constant over time, with noise affecting the elements of Y only. In particular, we assume that noise is additive, zero-mean, i.i.d. Gaussian with variance σ_n^2. In addition to noise, we assume that the main photometric distortions are due to illumination changes in the scene and variations of camera parameters (such as exposure and gain).

The main idea behind the novel change detection algorithm proposed in this paper is that photometric distortions occurring in the scene can be well modeled by means of the following *non-linear* parametric transformation:

$$y_i \;=\; a \cdot {x_i}^2 + b \cdot x_i + c\;, \qquad i = 1, \ldots, N \tag{1}$$

In particular, this class of transformations models non-linear variations occurring in real scenes by means of a quadratic polynomial function. It is important to note that this formulation is more general than the linear model [5, 6], which can not handle non-linearities. At the same time, the proposed model is less general than the order-preserving model [7, 8], for which any monotonic non-decreasing mapping function between the background and current frame intensities is considered as allowable and hence does not trigger the detection of a change.

For each pair of corresponding pixels in B and F, the proposed algorithm aims at computing the optimal Bayesian estimation of the three parameters $\vec{\Theta} = (a, b, c)$ given the two pixel sets X, Y. In particular, we aim at maximizing the *posteriori* probability given the *likelihood* and some *priors* on $\vec{\Theta}$.

Based on previous assumptions, the log-likelihood formulation of the Maximum-A-Posteriori (MAP) Bayesian estimation can be expressed as follows:

$$\vec{\Theta}_{MAP} = \left(\bar{a}, \bar{b}, \bar{c}\right)_{MAP} = \underset{(a,b,c)\in\mathbb{R}^3}{\operatorname{argmin}} \left(\frac{E\,(a,b,c)}{\sigma_n^2} + \frac{c^2}{\sigma_c^2} \right) \tag{2}$$

where $E\,(a, b, c)$ is given by:

$$E\,(a,b,c) = \sum_{i=1}^{N} \left(y_i - a x_i^2 - b x_i - c\right)^2 \tag{3}$$

and σ_c^2 is the variance of the zero-mean, Gaussian prior that we assume for c. More precisely, we hypothesize independence between the elements of $\vec{\Theta}$ and assume a,b as uniformly distributed and c as zero-mean Gaussian. The assumption on c is motivated by experimental evidence concerning the intensity mapping function observed in case of typical disturbance factors, which most often pass through the origin.

Denoted as λ the ratio between σ_n^2 and σ_c^2, the solution is obtained by equating to zero the partial derivatives with respect to the 3 parameters of the objective function in (2):

$$\frac{\partial\left(E(a,b,c) + \lambda c^2\right)}{\partial i} = 0, \qquad i = a, b, c \tag{4}$$

This yields the following linear system of normal equations:

$$\begin{bmatrix} Sx^4 & Sx^3 & Sx^2 \\ Sx^3 & Sx^2 & Sx \\ Sx^2 & Sx & N+\lambda \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} Sx^2y \\ Sxy \\ Sy \end{bmatrix} \tag{5}$$

where:

$$Sx = \sum_{i=1}^{N} x_i,\;\; Sx^2 = \sum_{i=1}^{N} x_i^2,\;\; Sx^3 = \sum_{i=1}^{N} x_i^3,\;\; Sx^4 = \sum_{i=1}^{N} x_i^4,$$
$$Sy = \sum_{i=1}^{N} y_i,\;\; Sxy = \sum_{i=1}^{N} x_i y_i,\;\; Sx^2y = \sum_{i=1}^{N} x_i^2 y_i$$

The system can be solved in closed form by Cramer's rule:

$$\bar{a} = \frac{D_a}{D}, \qquad \bar{b} = \frac{D_b}{D}, \qquad \bar{c} = \frac{D_c}{D} \tag{6}$$

where D is the determinant of the coefficient matrix in (5), and D_a, D_b, D_c are the determinants of the matrices obtained by replacing, respectively, the first, second and third column of the matrix with the vector at the right-hand side of (5).

Once $\bar{a}, \bar{b}, \bar{c}$ have been estimated as in (6), to obtain the final change mask, M, a pixel in F is classified as *C* (*Changed*) or *U* (*Unchanged*) based on the regression error between the data and the model predictions:

$$M = \begin{cases} C & \text{if } \; E\left(\bar{a}, \bar{b}, \bar{c}\right) + \lambda c^2 > \tau \\ U & \text{otherwise} \end{cases} \tag{7}$$

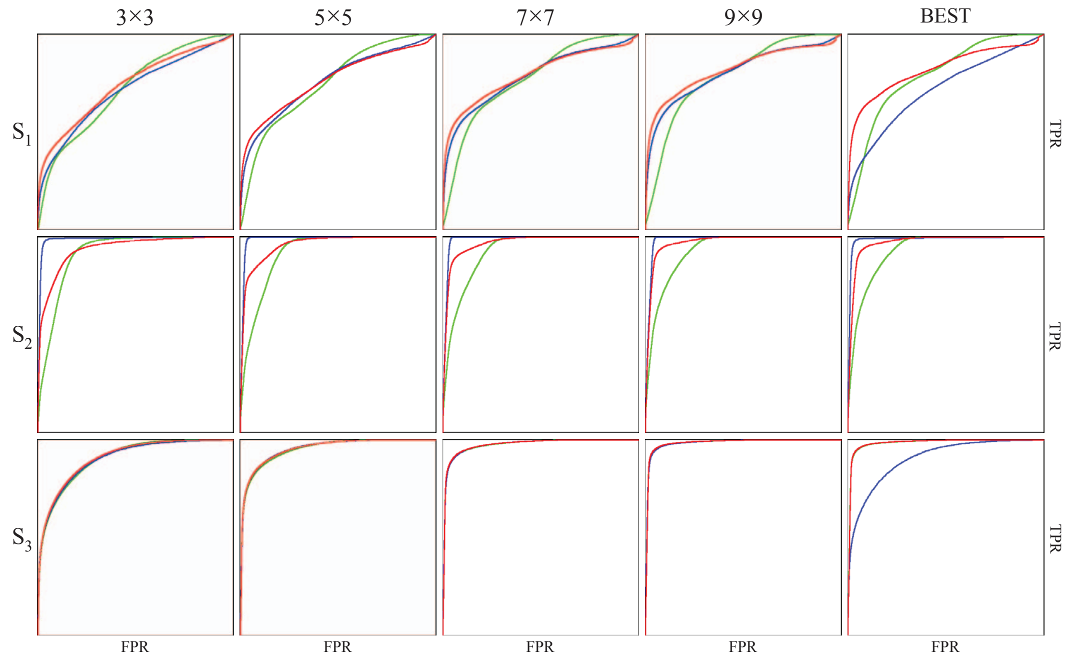
where τ is a fixed threshold.

By means of the proposed solution, we can compute the optimal estimation of the model parameters in a closed form, with no need to resort to any iterative approach. This renders our method particularly efficient. In addition, it is worth pointing out that all terms involved in the calculation of $\bar{a}, \bar{b}, \bar{c}$, i.e. those in the coefficient matrix and in the vector of constant terms of (5), can be computed either off-line (i.e. those involving only background pixels, Sx, Sx^2, Sx^3, Sx^4) or by means of very fast incremental techniques such as Summed Area Table [9] (i.e. Sy, Sxy, Sx^2y). Overall, this allows our method to exhibit a computational complexity of $O(1)$.

3. EXPERIMENTAL RESULTS

We present experimental results aimed at comparing the performance of our proposal to that of two state-of-the-art approaches that, analogously to ours, rely on modeling a priori the effect of disturbance factors over a neighborhood of pixel intensities. In particular, the considered algorithms, hereinafter referred to as O (Ohta) [6] and M (Mittal) [8], assume a linear and a non-parametric order-preserving model, respectively. We point out that we do not consider here algorithms belonging to the first class outlined in Section 1, such as [3, 4], since we are primarily interested in robustness against sudden illumination changes and, as discussed in Section 1, such algorithms are inherently unable to deal with these changes.

Experimental results have been obtained on three test sequences S_1, S_2, S_3 characterized by sudden and strong photometric changes due to disturbance factors. The background



FRAME RATE	O	M	P
3×3	111.7	6.1	38.8
5×5	106.7	1.3	37.9
7×7	107.6	0.3	36.9
9×9	106.7	0.1	36.8

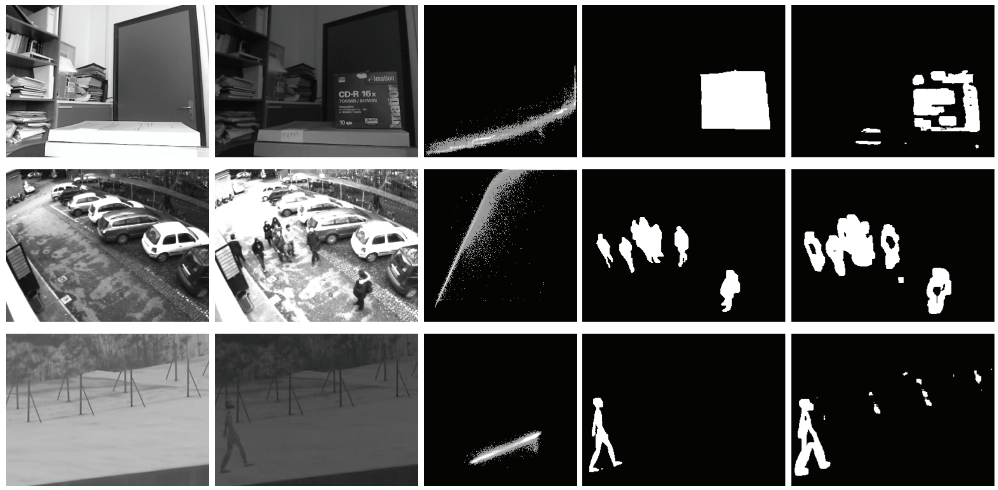
Fig. 1. Quantitative results: frame rates (left) and ROC curves reported by the evaluated algorithms.

and one sample frame for each sequence are shown in Figure 2. In particular, S_1 and S_2 are, respectively, a real indoor and outdoor sequence in which strong intensity variations occur due to both illumination changes and camera gain and exposure variations. S_3 is a synthetic sequence, available on the web for benchmarking purposes [10], in which moving people as well as photometric changes have been artificially superimposed on a real outdoor background. For each sequence, the background has been inferred off-line by averaging an initial sequence of frames free of moving objects. Then, algorithms have been run using square neighbourhoods of increasing size (3×3, 5×5, 7×7, 9×9).

As for computational efficiency, it is straightforward to derive that O has the same $O(1)$ complexity as the proposed algorithm, from now on referred to as P, while M is characterized by $O(N^2)$ complexity. In the table shown in Figure 1, left, we report a measure of the algorithms efficiency in terms of execution frame rate (in fps). The target PC is an Intel P4 3.06GHz. As expected, the frame rate of O and P does not depend significantly on the size of the neighbourhood while the frame rate of M quickly decreases with increasing sizes. In absolute terms, O and P are fast and allow for real-time processing with every choice of the neighbourhood size. On the contrary, only a 3×3 neighbourhood might be used to employ M in a real application.

As for accuracy, quantitative measures have been obtained, for each sequence, by computing the true positive rate (TPR) versus false positive rate (FPR) receiver operating characteristic (ROC) curve. Figure 1 shows the ROC curves obtained for each considered algorithm, sequence and neighbourhood size. Moreover, the rightmost column shows the *best* ROC curve achievable by the algorithms under the constraint of a reasonable processing speed with respect to typical application requirements. Hence, for O and P we report the ROC curve obtained with a 9×9 neighbourhood, while for M that obtained with a 3×3 neighbourhood.

The results in Figure 1 show that, overall, the proposed algorithm tends to achieve a better tradeoff between efficiency and robustness against disturbs compared to the other considered algorithms. In fact, though O is by far the most efficient algorithm, as can be seen from the frame rates table in Figure 1, unlike M and P it is not able to cope effectively with the typical disturbs occurring in real-world applications. As proof of that, we can note from the ROC curves that, for each neighbourhood size, O exhibits practically the same accuracy as M and P only in the synthetic sequence S_3 characterized by artificially superimposed photometric changes while in S_1 and S_2, where photometric variations are due to real disturbs, M and P outperform O. It is worth pointing out that, as shown by the background versus frame joint intensity histograms reported in the middle column of Figure 2, in the synthetic sequence S_3 the photometric changes are approximately linear (as hypothesized by O), while in the real sequences S_1 and S_2, disturbance factors yield non-linear transformations (i.e. the a-priori model adopted by M and P). As for those algorithms most suited to real-applications, Figure 1 shows that M outperforms P in S_2, and vice versa in S_1. However, the non-parametric order-consistency test deployed by M is much less efficient than the simple polynomial fitting procedure at the basis of our algorithm. Therefore, the



S_1

S_2

S_3

Fig. 2. Qualitative results yielded by the proposed algorithm on the 3 test sequences. For each sequence, from left to right: the inferred background, one sample frame, the joint intensity histogram, the ground truth mask and the binary change mask.

rightmost column of Figure 1 proves that, whenever the application calls for computational efficiency, the proposed algorithm turns out a better choice than M.

Finally, some qualitative results are presented in Figure 2. In particular, for each sequence it shows, from left to right, the inferred background, one sample frame, the background versus frame joint intensity histogram, the ground truth mask and the binary change mask yielded by the proposed algorithm. The change masks are obtained by using 9×9 neighbourhoods and by choosing, for each sequence, the threshold value that yields a FPR equal to 5%. These masks indicate notable robustness of the approach toward strong photometric changes yielded by disturbs affecting the considered frames, as can be judged by visually comparing corresponding background-frame pairs.

4. CONCLUSIONS

We have proposed a novel background subtraction approach relying on least-square quadratic polynomial fitting. Ability of the assumed polynomial transformation in modeling the local effect of disturbs, on one hand, and the simplicity of the fitting procedure, on the other hand, allow the proposed approach to outperform state-of-the-art methods in terms of robustness-efficiency tradeoff, as proved by experiments. Based on this promising results, we are currently investigating an alternative formulation for polynomial fitting which deploys an a-priori modeling of nuisances as a monotonic, homogeneous polynomial transformation.

5. 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 Sciences*, vol. 1, pp. 32–54, 2008.

[2] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, “Pfinder: Real-time tracking of the human body,” *Trans. PAMI*, vol. 19, no. 7, pp. 780–785, 1997.

[3] C. Stauffer and W. E. L. Grimson, “Adaptive background mixture models for real-time tracking,” in *Proc. CVPR*, vol. 2, 1999, pp. 246–252.

[4] A. M. Elgammal, D. Harwood, and L. S. Davis, “Non-parametric model for background subtraction,” in *Proc. ECCV*, vol. 2, 2000, pp. 751–767.

[5] E. Durucan and T. Ebrahimi, “Change detection and background extraction by linear algebra,” *Proc. IEEE*, vol. 89, no. 10, pp. 1368–1381, 2001.

[6] N. Ohta, “A statistical approach to background subtraction for surveillance systems,” in *Proc. ICCV*, vol. 2, 2001, pp. 481–486.

[7] B. Xie, V. Ramesh, and T. Boult, “Sudden illumination change detection using order consistency,” *Image and Vision Computing*, vol. 22, no. 2, pp. 117–125, 2004.

[8] A. Mittal and V. Ramesh, “An intensity-augmented ordinal measure for visual correspondence,” in *Proc. CVPR*, vol. 1, 2006, pp. 849–856.

[9] F. Crow, “Summed-area tables for texture mapping,” *Computer Graphics*, vol. 18, no. 3, pp. 207–212, 1984.

[10] MUSCLE Network of Excellence. Motion detection video sequences. Pattern Recognition and Image Processing Group, Vienna University of Technology. [Online]. Available: http://muscle.prip.tuwien.ac.at/