Signal Processing 89 (2009) 1412–1416

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

# Continuous variance estimation in video surveillance sequences with high illumination changes

Pedro Gil-Jiménez *, Hilario Gómez-Moreno, Javier Acevedo-Rodríguez, Saturnino Maldonado Bascón

*Dpto. de Teoría de la señal y Comunicaciones, Universidad de Alcalá, 28805 Alcalá de Henares, Madrid, Spain*

ARTICLE INFO

*Article history:*
Received 24 November 2008
Received in revised form
16 January 2009
Accepted 19 January 2009
Available online 30 January 2009

*Keywords:*
Video surveillance
Motion detection
Continuous variance estimation

ABSTRACT

Continuous estimation of signal statistics is an important issue in many video processing systems, such as motion detection in surveillance applications. In this paper we demonstrate how results of classical expressions for variance estimation decrease in accuracy when dealing with sequences containing high illumination variations. The paper also proposes a new estimation method, and shows how, under such conditions, the accuracy of the proposed method produces better results whilst maintaining performance in scenarios with smaller changes, thus improving the motion detection stage of a video surveillance system.



## 1. Introduction

Background maintenance is a common block in video surveillance systems. It enables the system to update the background model, that is, some of the background statistics [1]. Frequently, the statistics used are the mean, which is an accurate representation of the background, and the variance, which provides information about the behavior of each zone of the image. For motion detection, both parameters are used together to determine whether a pixel corresponds to the background or foreground. For instance, in [2], if a pixel $p(x,y)$ has a value which is more than twice its typical deviation from the mean, then that pixel is considered as belonging to the foreground, that is:

$$M(x,y)=\begin{cases}\text{Background} & \text{if } \mu_{xy}-2\sigma_{xy}<p(x,y)<\mu_{xy}+2\sigma_{xy}\\ \text{Foreground} & \text{otherwise}\end{cases} \tag{1}$$

According to this, the greater the variance of a given pixel, the lower is the sensitivity, since the range of possible values for belonging to the foreground decreases. In [3], the variance is further used, along with other parameters, to classify each zone of the image according to their behavior of each zone. In this case, an error in variance computation would lead to a pixel misclassification.

If we consider a video sequence, such as the one shown in Fig. 1(a), several statistics can be computed. Fig. 1(b) shows the estimated mean. As we can see, this image depicts the background of the scene, that is, an image of the scene with all moving objects removed. The variance of the sequence can also be computed, in this case shown in Fig. 2(a) or (b). Note that since the variance is not itself an image, it must be normalized in order to be visualized properly in a figure. In this case, black (pixel = 0) corresponds to a variance equal to 0, and, linearly, white (pixel = 255) corresponds to a variance equal to 1000. In this image, we can see how pixels with high variance values (white levels in the normalized image) belong to high activity zones, for instance, the road, and low values to low activity ones.

Although many more statistic can be computed, these are the two basic ones proposed in most works. However, since the input of a video surveillance system is a stream of unlimited length, these statistics need to be computed

* Corresponding author.
*E-mail address:* pedro.gil@uah.es (P. Gil-Jiménez).

0165-1684/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2009.01.013


**Fig. 1.** (a) One image of a sequence and (b) its estimated background.

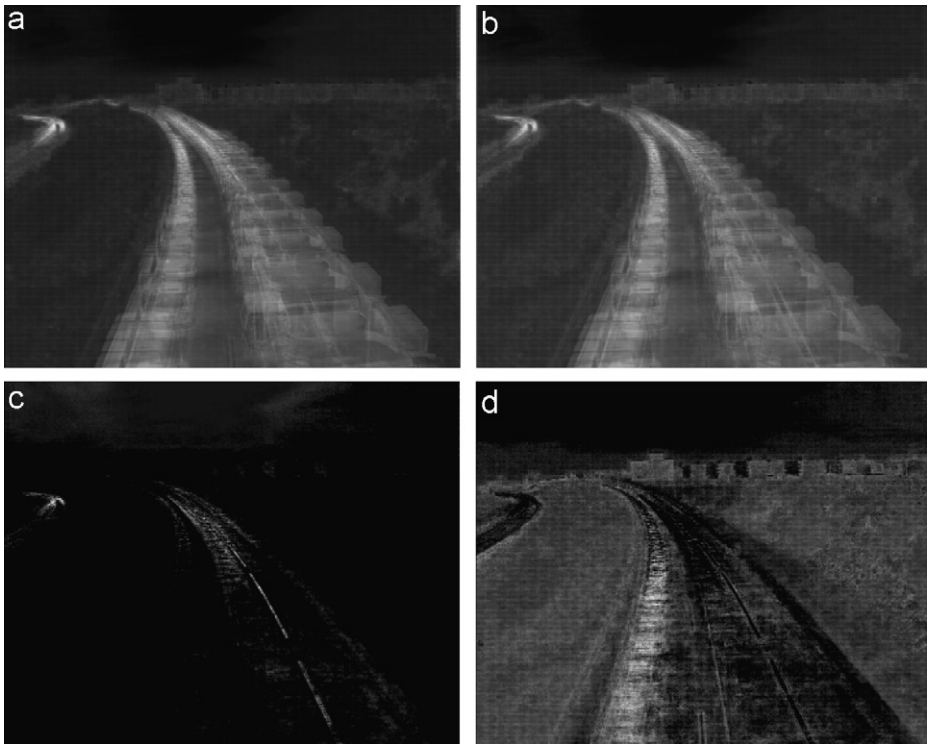


**Fig. 2.** Normalized estimated variance of the video sequence of Fig. 1 using (a) expression (8) and (b) expression (3). In both figures, black pixels correspond to variance values equal to 0, whereas white pixels to 1000. (c) (a) minus (b) where positive, 0 otherwise. (d) (b) minus (a) where positive, 0 otherwise.

continuously, that is, their values must be updated with each new frame. For that reason, results shown in Figs. 1(b) and (2) are the statistical estimations for one particular moment in the sequence. To this end, several expressions have been proposed in the literature; for estimation of the mean, one of the best known is the one employed in [4–6]:

$$\overline{\mu}_t = (1-\alpha)\overline{\mu}_{t-1} + \alpha I_t \quad (0<\alpha<1) \tag{2}$$

where subscript $t$ refers to frame number, $I_t$ is the image itself, $\overline{\mu}_t$ is the estimated mean and $\alpha$ is called the learning rate. Since we are considering images, (2) has to be computed independently for each pixel.

## 2. Variance estimation

Whilst (2) is one of the preferred expressions for the estimation of the mean, several others have been proposed in the literature for variance. For instance [7] proposes the following estimator, $\overline{\sigma}_t^2$:

$$\overline{\sigma}_t^2 = (1-\alpha)\overline{\sigma}_{t-1}^2 + \alpha(I_t - \overline{\mu}_t)^2 \tag{3}$$

Other expressions have also been defined in the literature [2,8,9], but their performance remains basically the same. Nonetheless, variance can be estimated in many other ways. For instance, in [10], an initial estimation of variance is computed in a similar fashion to above, then its histogram is computed, and finally the definitive value is obtained from the first peak in that histogram. In [4,11], variance is estimated in the same way as in (3), although both propose modeling the background as a mixture of Gaussians, and the expression is used to update only the current Gaussian.

Although (3) has proven to be accurate enough when working with real images, in the case of sequences with high illumination variations, (3) shows a systematic error. An analysis can be carried out of the expected value of the last estimator in order to demonstrate this error. In the case of a signal without illumination changes, where the mean value of the signal does not vary and consequently (2) can track its value accurately, that is $\mathrm{E}\{\overline{\mu}_t\} = \mu$ where $\mu$ is the actual mean of a given pixel, the expected value for (3) is

$$\mathrm{E}\{\overline{\sigma}_t^2\} = (1-\alpha)\mathrm{E}\{\overline{\sigma}_{t-1}^2\} + \alpha\mathrm{E}\{(I_t-\mu)^2\} \tag{4}$$

Taking into account that, by definition, $\mathrm{E}\{(I_t-\mu)^2\} = \sigma^2$:

$$\mathrm{E}\{\overline{\sigma}_t^2\} = (1-\alpha)\mathrm{E}\{\overline{\sigma}_{t-1}^2\} + \alpha\sigma^2 = \sigma^2 \tag{5}$$

that is, (3) is a good estimation when the illumination does not vary at all, or at least when variations occur very slowly. However, in real scenarios, illumination conditions can cause pixel values to vary rapidly so that the estimated mean differs from its actual value in $\Delta_t = \overline{\mu}_t - \mu$. For the sake of simplicity, and supposing that $\Delta_t$ is constant, i.e., $\Delta_t = \Delta$, the expected value for the second term will be

$$\begin{aligned}&\mathrm{E}\{(I_t-\overline{\mu}_t)^2\}\\ &\quad= \mathrm{E}\{(I_t-(\mu+\Delta))^2\}\\ &\quad= \mathrm{E}\{I_t^2 - 2I_t\mu + \mu^2 + \Delta^2 - 2I_t\Delta + 2\mu\Delta\}\\ &\quad= \mathrm{E}\{I_t^2 - 2I_t\mu + \mu^2\} + \Delta^2 - 2\Delta\mathrm{E}\{I_t\} + 2\mu\Delta\end{aligned} \tag{6}$$

The first term corresponds to signal variance, whereas the last two terms cancel each other out; thus, the expected value is

$$\mathrm{E}\{(I_t-\overline{\mu}_t)^2\} = \sigma^2 + \Delta^2 \tag{7}$$

that is, the estimator has a systematic error with a value of $\Delta^2$, and this error increases as the difference between the estimated mean $\overline{\mu}_t$ and the actual mean $\mu$ increases. With the aim of improving estimation accuracy in cases of high illumination variability, in this paper we propose the modification of (3) to the following expression:

$$\overline{\sigma}_t^2 = (1-\alpha)\overline{\sigma}_{t-1}^2 + \frac{\alpha}{2}(I_t - I_{t-1})^2 \tag{8}$$

Leaving aside for the moment the term $1/2$, the validity of the expression can be checked. The expected value for the second term will be

$$\mathrm{E}\{(I_t - I_{t-1})^2\} = \mathrm{E}\{I_t^2\} + \mathrm{E}\{I_{t-1}^2\} - 2\mathrm{E}\{I_tI_{t-1}\} \tag{9}$$

for the last term, when the samples are uncorrelated, as may be the case for some pixels, especially those belonging to noisy zones of the scene such as trees, road, water, etc., we have

$$\mathrm{E}\{I_tI_{t-1}\} = \mathrm{E}\{I_t\}\mathrm{E}\{I_{t-1}\} = \mu^2 \tag{10}$$

and thus,

$$\mathrm{E}\{(I_t - I_{t-1})^2\} = 2(m_2 - \mu^2) = 2\sigma^2 \tag{11}$$

which demonstrates that, for the noisy pixels in the image, the variance computed with (8) yields a more accurate estimation than that computed with (3). Correlated pixels usually belong to a static background, where variance is near zero and, although (10) does not hold in such cases, the proposed expression gives a value near zero which, for the video surveillance problem discussed here, is absolutely valid. In any case, the goal of this stage is not to provide an accurate estimation of variance in all possible cases, but rather to give a useful estimation for further stages, especially motion detection, as is the case with the method proposed in this paper.

## 3. Experimental results

An example can help to better understand the improvement achieved with the proposed expression. Let us suppose a random signal which corresponds to a pixel with variance equal to 1, and a mean value equal to 0 constant over time until one particular instant, where the mean value changes suddenly to a value equal to 2, as can be seen in Fig. 3. This signal constitutes an approximated model of the illumination changes typical of certain video surveillance scenarios.

Fig. 4 shows the evolution of signal variance computed using expressions (3) and (8). Apart from the transient period, where both signals need to reach their permanent value, it can be seen that variance estimated with the proposed expression remains close to a value of 1 throughout the whole sequence. This represents the actual value of the variance, whereas the variance estimated with (3) produces a very different value when the illumination change occurs, and this deviation is repeated in the following samples until that value stabilizes.

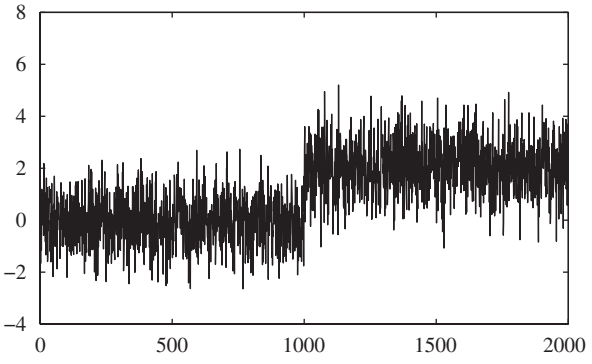


**Fig. 3.** Non-stationary signal with variance equal to 1.

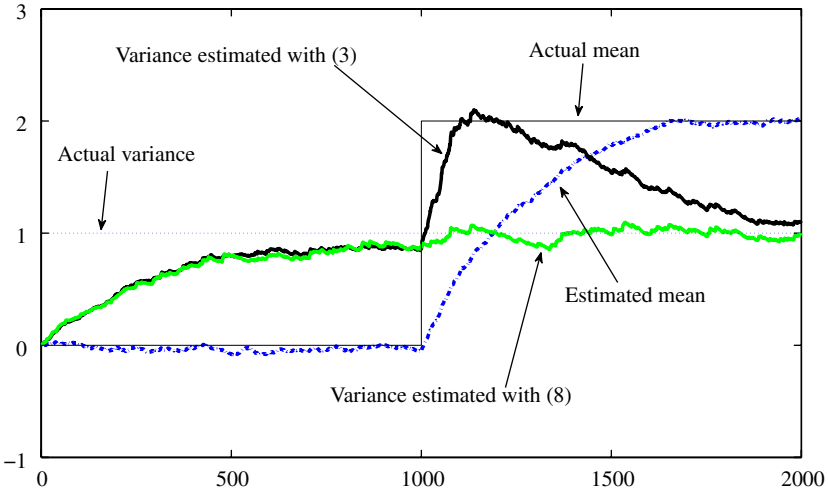


**Fig. 4.** Estimated variance for the signal in Fig. 3.

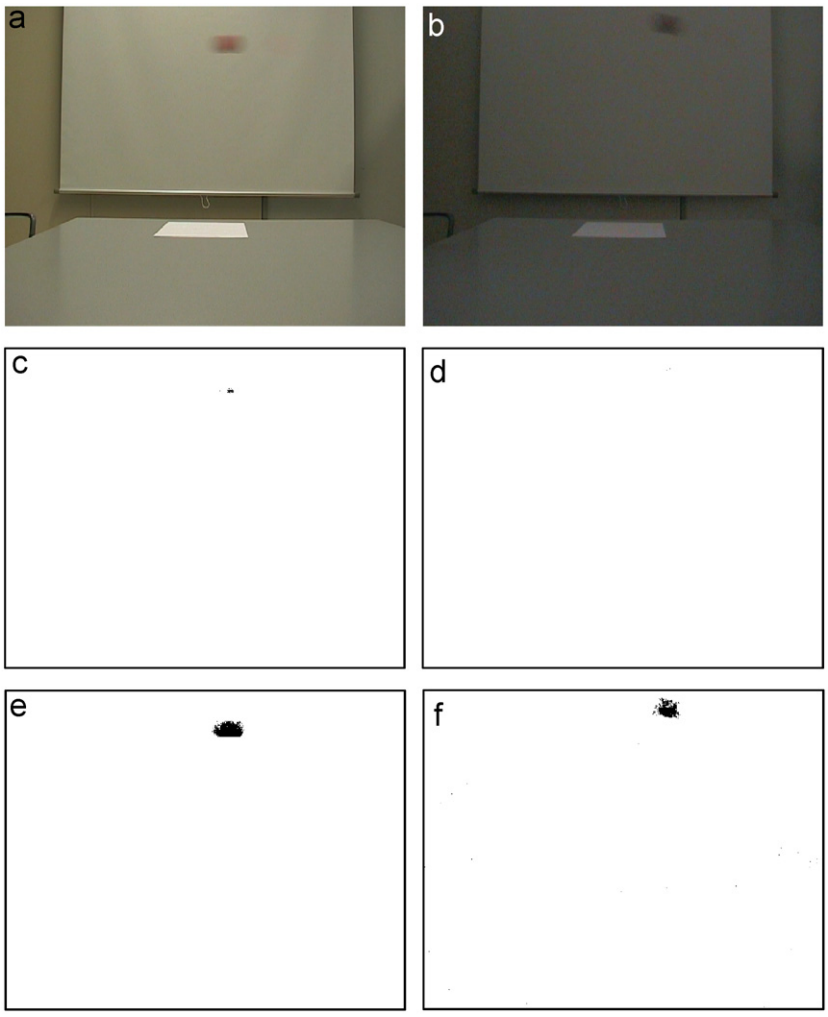


**Fig. 5.** Motion detection masks. (a) and (b) Two different images of the sequence. (c) and (d) Motion masks obtained from the variance computed with expression (3). (e) and (f) Motion masks obtained from the variance computed with expression (8).

Lastly, we can check the performance of both expressions working in a real scenario. As we saw before, Fig. 2(a) and (b) shows the estimated variance, at one particular moment, in the video sequence shown in Fig. 1, using expression (8) (Fig. 2(a)), and expression (3) (Fig. 2(b)). These results were obtained from a few frames taken after an illumination change, caused by passing clouds. The zone of interest for these results is focused on those parts of the image corresponding to the grass and trees (bottom half of the figure, except the road). Although variance in these zones is much lower than that of the road, and remains so for both expressions throughout the beginning of the sequence, we can see that, after the illumination change, variance for these zones rises when computed using expression (3), reaching a value almost equal to the variance of the road, which is, obviously, incorrect. This difference can be better appreciated in Fig. 2(c) and (d). Fig. 2(c) shows the value of the difference between figures (a) and (b), magnified 5 times when this value is positive, and black for the pixels where the difference is negative, that is

$$d(x,y)=\begin{cases}5(\bar{\sigma}^2_{xy(3)}-\bar{\sigma}^2_{xy(8)}) & \text{if } \bar{\sigma}^2_{xy(3)}-\bar{\sigma}^2_{xy(8)}>0\\ 0 & \text{otherwise}\end{cases} \quad (12)$$

where $\bar{\sigma}^2_{(3)}$ refers to the variance computed with (3) and so on. Fig. 2(d) represents exactly the opposite. As we can see, for the zones of interest, the variance computed with (3) is always greater than that computed with (8), as has been demonstrated above.

The consequence of this drift can be analyzed from the point of view of the motion detection block. As we saw before, motion detection systems using an expression similar to (1) undergo some loss of sensitivity when computed variance is greater than real variance. This can be appreciated in the experiment shown in Fig. 5. In this figure, (a) and (b) correspond with two particular moments of an inner sequence. This sequence was recorded as follow. At the beginning of the sequence, the scenario had low illumination. After some time, the room lights were switched on, to produce a swift illumination change, and some seconds afterwards, an object crossed the scene. This particular instant is shown in Fig. 5(a). Subsequently, the room lights were switched off, to produce a new illumination change, and again, an object crossed the scene few seconds afterwards, as it is represented in Fig. 5(b). If we extract the motion detection in the scene using (1) and the variances computed with both expressions, we get the mask shown in Fig. 5(c) and (d) for expression (3), and (e) and (f) for (8). As we see before, since the variance computed with (3) is always greater than that computed with (8), there is a loss of sensitivity in the first case, as can be seen in (c), where the size of the mask corresponding to the detected object is much smaller than the one in (e), or in (d), where the object is not detected at all.

We have to remark that the improvement achieved in this paper is specially important for those zones of the image where the variance is not so high. When variance is so high, normally due to non-static background as it is the case of a road with vehicles in motion, that there is a complete loss of sensitivity in that part of the scene, other motion detection strategies must be employed, and these are not normally affected by an error in the variance computation.

## 4. Conclusions

A new method for the continuous estimation of video sequence variance has been proposed in this paper. It yields better results than standard methods when analyzing noisy pixels in sequences with high illumination variability, which is frequently the case in real, uncontrolled scenarios, and yields similar results in sequences without illumination changes. An experiment has also been carried out in order to demonstrate these results in practical applications. The conclusions obtained in this paper could help to improve the performance of video surveillance systems working in uncontrolled environments, such as outdoor scenarios.

## Acknowledgments

This work was supported by Comunidad de Madrid-UAH, Project no. CCG07-UAH/TIC1740 and Ministerio de Ciencia e Innovación Project no. TEC2008-0277/TEC.

## References

[1] K. Toyama, J. Krumm, B. Brumitt, B. Meyers, Wallflower: principles and practice of background maintenance, in: ICCV, vol. 1, 1999, pp. 255–261.
[2] T. Kanade, R.T. Collons, A.J. Lipton, Advances in cooperative multi-sensor video surveillance, in: DARPA Image Understanding Workshop, 1998, pp. 3–24.
[3] P. Gil-Jiménez, S. Maldonado-Bascón, R. Gil-Pita, H. Gómez-Moreno, Background pixel classification for motion detection in video image sequences, in: Computational Methods in Neural Modeling, Lecture Notes in Computer Science, vol. 2686, Springer, Berlin, 2003, pp. 718–725.
[4] C. Stauffer, W. Eric, L. Grimson, Adaptive background mixture models for real-time tracking, in: CVPR, 1999, pp. 2246–2252.
[5] D. Koller, J. Weber, J. Malik, Robust multiple car tracking with occlusion reasoning, in: ECCV, vol. 1, 1994, pp. 189–196.
[6] K.S. Bhat, M. Saptharishi, P.K. Khosla, Motion detection and segmentation using image mosaics, in: IEEE International Conference on Multimedia and Expo, vol. III, 2000, pp. 1577–1580.
[7] T. Boult, R. Micheals, X. Gao, M. Eckmann, Into the woods: visual surveillance of noncooperative and camouflaged targets in complex outdoor settings, in: Proceedings of the IEEE, vol. 89, 2001, pp. 1382–1402.
[8] H. Fujiyoshi, A. Lipton, Real-time human motion analysis by image skeletonization, in: IEEE Workshop on Applications of Computer Vision, 1998.
[9] S.J. McKenna, S.J.Z. Duric, A. Rosenfeld, H. Wechsler, Tracking groups of people, Computer Vision and Image Understanding 80 (2000) 42–56.
[10] J.C. Silveira-Jacques, C. Rosito-Jung, S. Raupp-Musee, A background subtraction model adapted to illumination changes, in: Proceedings of the IEEE ICIP, ICIP'06, 2006, pp. 1817–1820.
[11] Y.-L. Tian, M. Lu, A. Hampapur, Robust and efficient foreground analysis for real-time video surveillance, in: Proceedings of the 2005 IEEE CVPR, 2005.