

# ESTIMATION OF DYNAMIC BACKGROUND FOR FLUORESCENCE VIDEO-MICROSCOPY\*

*J. Boulanger, C. Kervrann*

IRISA–INRIA / INRA–MIA  
Campus Universitaire de Beaulieu  
35042 Rennes Cedex - France

*P. Bouthemy*

IRISA–INRIA  
Campus Universitaire de Beaulieu  
35042 Rennes Cedex - France

## ABSTRACT

This paper describes a method for separating moving objects from temporally varying background in time-lapse confocal microscopy image sequences representing fluorescently tagged moving vesicles. A temporal linear model is considered for background modeling whose parameters are robustly estimated using asymmetric M-estimators combined with a bias-variance trade-off criterion. Furthermore, we propose an original approach for automatically detecting moving objects in the image sequence. Experimental results demonstrate the interest of this proposed method which can be relevant for biological studies from image sequences.

## 1. INTRODUCTION

Time-lapse confocal microscopy is now a popular technology in biomedicine and biophysics for the analysis of small moving objects at sub-cellular or molecular levels. Several applications in biology are concerned with the acquisition of sequences of  $2D$  images or  $3D$  volumes representing small fluorescently tagged particles with highly variable velocities. Objects of interest can be vesicles moving in the cell medium. Even if the medium is not directly tagged with fluorescent proteins, (undesirable) large structures belonging to the background are still visible in the image sequence. Accordingly, several tasks such as detection or tracking of moving vesicles require to distinguish the tagged particles from the background.

In video image sequence analysis, the detection of moving objects is usually easier if a background subtraction procedure is applied [1, 2]. However, in the case of images showing fluorescently tagged particles, the global image intensity can vary slowly along time. This can be due to several physical phenomena such as photo-bleaching or diffusion of fluorescent proteins within the cell. Therefore a stationary model for the background is too restrictive and the moving particles (e.g., vesicles) would not be successfully detected. However,

we have conducted some experiments showing that the intensity variation w.r.t. time can be captured by a linear model for each pixel, since we are dealing with sequences of limited length. This modeling provides a compact representation of a background intensity dynamics. It can be described by two maps representing the parameters of the linear model for each spatial location. In addition, the involved parameters are spatially correlated and this must be taken into account in the estimation process. Let us mention that the proposed method could be applied to non linear intensity models as well.

In contrast to Bayesian methods our method is local and follows a point-wise adaptive estimation framework [3, 4]. We do not either model the objects as usually done, but we only assume that they correspond to bright spots against the background.

The remainder of the paper is organized as follows. In Section 2 the image model is introduced. The proposed estimation method is then described in two steps. In Section 3, we first address the problem of estimating background linear model considering separately temporal signals. An adaptive asymmetric estimator is proposed especially in order to take into account the form observed signal. In Section 4, the spatial coherence of the temporal linear models is introduced into the estimation process through the bias-variance trade-off criterion. In Section 5, we present how we select the points belonging to the background. In Section 6, we report experimental results which demonstrate that we can separate the moving vesicles from the slowly varying background. This method has been applied to the analysis of the influence of the Rab6 proteins on the intra-cellular traffic. Fluorescently tagged proteins allow the localization of vesicles propelled by motor proteins along with a network of microtubules.

## 2. IMAGE MODEL

The following image sequence model is considered

$$f(x, y, t) = a(x, y) + b(x, y)t + u(x, y, t) + \epsilon(x, y, t) \quad (1)$$

where  $f(x, y, t)$  denotes the intensity observed at pixel  $(x, y)$  and time  $t$ , the two coefficients  $a(x, y)$  and  $b(x, y)$  varies with the spatial image position  $p_i = (x, y)$ .  $u(x, y, t)$  is a positive

---

\*THE AUTHOR WOULD LIKE TO THANK THE ACI-IMPBIO (MO-DYNCELL5D. PROJECT) FOR ITS SUPPORT AND CURIE INSTITUTE FOR PROVIDING IMAGE SEQUENCES

function that describes the intensity of moving vesicles and  $\epsilon(x, y, t)$  is an additive white Gaussian noise. In the sequel, we will use the subscript  $i$  to denote the spatial position  $p_i$ , and accordingly (1) can be stated as

$$f_i(t) = a_i + b_i t + u_i(t) + \epsilon_i(t). \quad (2)$$

### 3. PIXEL-WISE ESTIMATION OF THE BACKGROUND MODEL PARAMETERS

We first deal with the estimation of parameters  $a_i$  and  $b_i$  for a single temporal 1D signal. Let us point out that that this estimation has to be performed several millions of times (for each image point). Accordingly, the estimation procedure must be very fast. Moreover, vesicles have an erratic behavior and sometimes stop for a long time. Then, prior motion detection can not be used here to extract the objects from the background. Our background estimation will then be based on the image intensity only. Since we must not be affected by the presence of moving vesicles, we will resort to a robust estimation framework.

#### 3.1. Robust estimation

The two parameters  $a_i$  and  $b_i$  are estimated by minimizing a robust error function:

$$E(a_i, b_i) = \sum_{t=1}^n \rho(f_i(t) - (a_i + b_i t)), \quad (3)$$

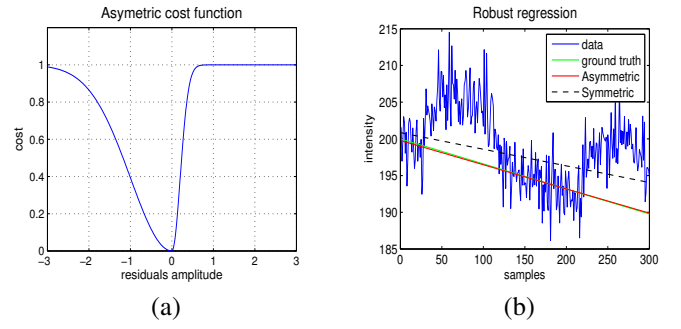
where  $n$  is the number of signal samples,  $r_i(t) = f_i(t) - (a_i + b_i t)$  is the residuals and  $\rho(\cdot)$  is a robust norm function. A local minimum is obtained by using the iterative re-weighted least squares (IRLS) procedure.

The choice of the robust function  $\rho$  is usually driven by the noise probability density function [5]. In our case, the overall noise is the sum of two components:  $u_i(t)$  and  $\epsilon_i(t)$  as defined in Section 2. To take into account that  $u_i(t)$  usually takes high positives values (vesicles appear as bright spots in the images), we adopt an asymmetric robust function (Leclerc estimator [6, 7]) plotted in Figure 1a and defined by:

$$\rho(z) = \begin{cases} 1 - \exp(z^2/\lambda\sigma_1) & \text{if } x \leq 0, \\ 1 - \exp(z^2/\lambda\sigma_2) & \text{otherwise.} \end{cases} \quad (4)$$

The scale  $\sigma_2$  factor can be estimated by applying a robust least trimmed squares (LTS) estimator to the pseudo-residuals defined as [8]:  $s_i(t) = (f_i(t+1) - f_i(t))/\sqrt{2}$ , where the coefficient  $1/\sqrt{2}$  insures that  $\mathbb{E}[s_i(t)^2] = \mathbb{E}[f_i(t)^2]$ . The scale factor  $\sigma_1$  is estimated by using the variance of the residuals given by the least mean squares estimator obtained at the initialization. We just point out that, in region where there is no moving vesicles,  $\sigma_1$  and  $\sigma_2$  are found almost equal. Finally,  $\lambda$  is chosen in the range [1, 3].

As a matter of fact, the proposed estimator is biased [7] but the bias is small and simulations proved that the  $L_2$  risk is



**Fig. 1.** (a) Asymmetric Leclerc robust function. (b) Regression using an asymmetric robust function in the case of a realistic simulation. Comparison with a symmetric Leclerc function. The asymmetric estimator (red) fits perfectly the ground truth (green) while the symmetric function provides biased result (dotted line).

smaller when an asymmetric cost function is used and when the data are corrupted by an additive positive signal. Figure 1b shows that the proposed estimator is able to deal with heavily contaminated data and outperforms the symmetric Leclerc M-estimator.

#### 3.2. Confidence matrix

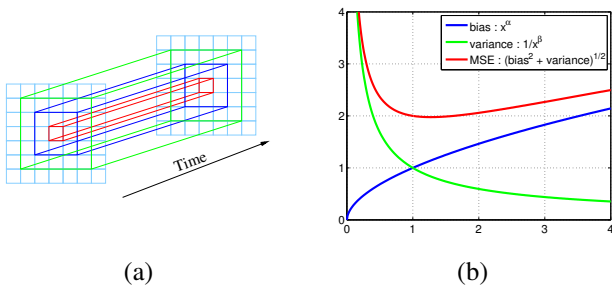
An accurate estimation of the confidence matrix for the estimated parameters is needed for the subsequent steps described in Section 4. We use the approximation proposed in [5] to compute the estimation covariance matrix:

$$\hat{C}_i = \frac{\sum_{t=1}^n w_{i,t}(r_i(t))r_i(t)^2 \sum_{t=1}^n w_{i,t}^2(r_i(t))}{\left(\sum_{t=1}^n w_{i,t}(r_i(t))\right)^2} \times \left( \begin{array}{cc} \sum_{t=1}^n w_{i,t}(r_i(t)) & \sum_{t=1}^n w_{i,t}(r_i(t))t \\ \sum_{t=1}^n w_{i,t}(r_i(t))t & \sum_{t=1}^n w_{i,t}(r_i(t))t^2 \end{array} \right)^{-1} \quad (5)$$

where  $r_i(t) = f_i(t) - (a_i + b_i t)$  and the weights are given by  $w(z) = \rho'(z)/z$ . Unlike the expression given in [9], the approximation given by Eq (5) is not asymptotic and yields a better estimation of the covariance matrix when  $n$  is small.

### 4. ESTIMATION OF THE BACKGROUND MODEL PARAMETERS WITH SPATIAL COHERENCE

We now introduce spatial coherence to design a regularized estimation of the parameters  $a_i$  and  $b_i$ . This can be accomplished by adopting the bias-variance trade-off framework [3, 10, 11, 4]. Instead of using just one temporal intensity signal at pixel  $p_i$  to estimate the parameters of the background intensity model as described in Section 3, a set of temporal 1D signal, collected in a neighborhood of the pixel  $p_i$  is now exploited in order to take into account the spatial coherence of the parameters. We will define a content adaptive specification of the neighborhood considered at each point  $p_i$ . For this purpose, a set of nested space-time tubes is introduced



**Fig. 2.** (a) Set of nested tube  $\{\mathcal{T}_{i,k}\}_{k=1,\dots,3}$ . (b) Bias-variance trade-off principle. When the diameter of the tube increases, the bias increases and the variance decreases. The optimum is achieved when the bias and the variance are of the same order.

as plotted in Figure 2(a), by taking the pixels in a growing square spatial neighborhood of  $p_i$ . Each tube  $\mathcal{T}_{i,k}$  at  $p_i$  can be parametrized by its diameter  $\phi_{i,k}$ ,  $k \in [1, \dots, K]$ :

$$\mathcal{T}_{i,n} = \{f(x_j, y_j, z_j) : |x_i - x_j| + |y_i - y_j| < \phi_{i,k}\} \quad (6)$$

In order to select the optimal diameter of the space-time tube, we propose to minimize the point-wise  $L_2$  risk of the parametric estimator defined as  $\mathbb{E}[(\hat{\theta}_i - \theta_i)^2]$  where  $\theta_i = (a_i, b_i)$  is the true parameter pair and  $\hat{\theta}_i$  its corresponding estimator, at position  $p_i$ . The  $L_2$  risk can be decomposed into two parts: squared bias and variance. As shown in Figure 2b, while the diameter  $\phi_{i,k}$  increases with  $k$ , the bias increases too. This can be explained by the fact that, the data cannot be described any longer by a unique model. By taking more and more data points, the variance decreases. This behavior, also called bias-variance trade-off [3], is exploited to detect the minimum of the point-wise  $L_2$  risk which is equal to the sum of square mean and variance.

For each diameter  $\phi_{i,k}$ , new estimates of the background model parameters  $\hat{\theta}_{i,k}$  and the associated covariance matrix  $\hat{C}_{i,k}$  are computed with the same procedure as the one described in Section 3 but using all the data from the considered neighborhood. It can be shown that the bias variance trade-off can be expressed with the following test [4] (here a vectorial version is given):

$$\frac{n-2+1}{2n} (\hat{\theta}_{i,k} - \hat{\theta}_{i,k'})^\top \hat{C}_{i,k'}^{-1} (\hat{\theta}_{i,k} - \hat{\theta}_{i,k'}) < \eta \quad (7)$$

for every  $0 \leq k' < k$ , the diameter of the tube is increased and the estimation process is continued. It can be demonstrated that the threshold  $\eta$  can be defined as a quantile of a Fisher distribution of parameters 2 and  $n-2-1$ .

## 5. BACKGROUND/OBJECT SEPARATION

In this section, we propose a method to decide which pixels belong to the background. It can be considered as an indirect but effective way to detect the objects of interest in the

image sequence. We follow the approach described in [12] and which was applied the differential analysis of micro-array data. It is based on a penalized likelihood criterion that aims at estimating the mean and the variance of a Gaussian vector assuming some of the components are zero. We suppose that the number of zero components as well and their positions are unknown. In our problem, the zero components are associated to the background while the other components represent the objects. Formally, we have:

$$\mathbf{Z} = \mathbf{m} + \mathbf{e}, \quad \mathbf{e} \sim \mathcal{N}(0, \Sigma), \quad (8)$$

where  $\mathbf{Z}$  is the vector of  $n$  residuals defined by  $z(t) = f_i(t) - (a_i + b_i t)$ , assumed to be independent, and  $\mathbf{e}$  is a Gaussian white noise of variance  $\tau^2 I_n$ . The expectation  $\mathbb{E}[\mathbf{Z}] = \mathbf{m} = (m_1, \dots, m_n)^\top$  and the variance  $\tau^2$  are unknown. Moreover,  $k_0$  component of  $\mathbf{m}$  are assumed to be non zero. The data are re-ordered such that  $|z_j| > |z_{j+1}|$  and the following model collection  $\mathcal{M} = \{M_1, \dots, M_n\}$  is considered:

$$\begin{cases} M_1 = (z_1, 0, \dots, 0)^\top \\ M_2 = (z_1, z_2, 0, \dots, 0)^\top \\ M_n = (z_1, z_2, z_3, \dots, z_n)^\top \end{cases} \quad (9)$$

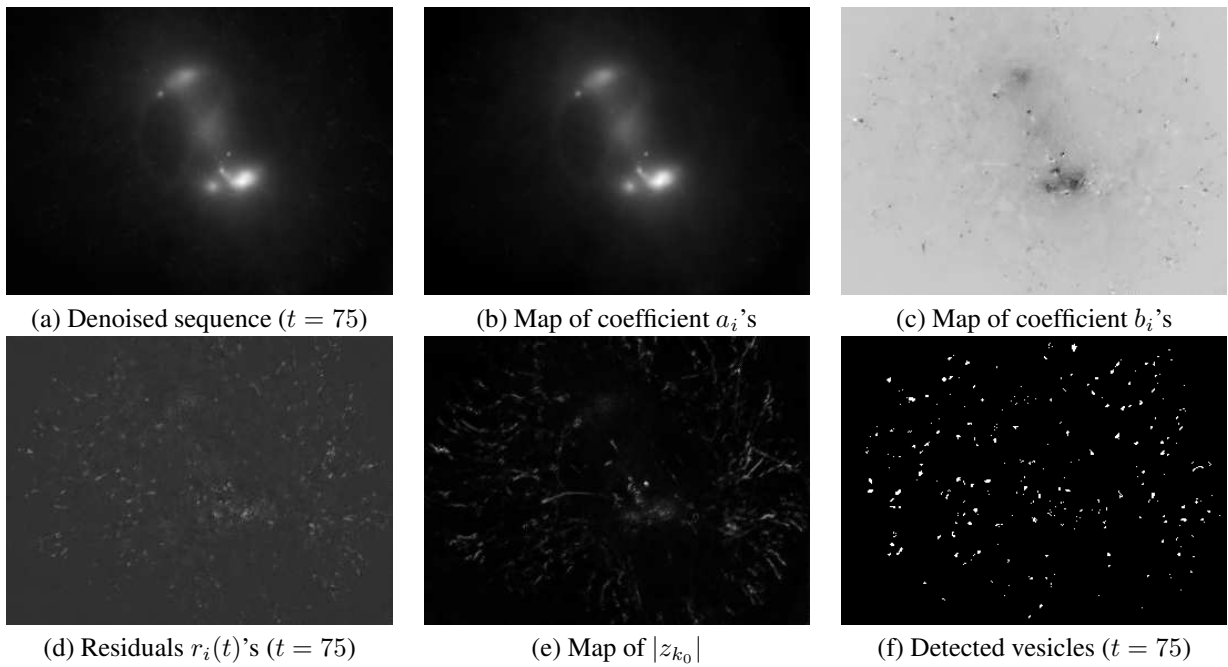
The selected model is the one minimizing the following penalized likelihood criterion [13]:

$$J(M_j) = \frac{n}{2} \log(\hat{\tau}_j^2) + n \left( c_1 \log\left(\frac{n}{k_j}\right) + c_2 \right) \frac{k_j}{n - k_j} \quad (10)$$

where  $\hat{\tau}_j^2 = (\sum_{k=j}^n z_k^2)/n$  is the plugged-in maximum likelihood estimator of  $\tau_j^2$ . The two universal constants  $c_1$  and  $c_2$  were calibrated in [12] and found as  $c_1 = 2$  and  $c_2 = 4$ . This method provides the set of points belonging to the background, and by taking the complementary set, we have a simple mean to detect the vesicles.

## 6. EXPERIMENTAL RESULTS

The described method has been applied to real image sequences representing small GFP-tagged vesicles moving into the cell. As a preprocessing step, a denoising procedure has been first applied to the image sequences (Fig. 3a). The estimated background is then compactly represented by three maps. The first one shown on Fig. 3b corresponds to the map of coefficients  $a_i$ 's of the model. It represents the shape of the background at  $t = 0$ . The second one is the map of the linear coefficients  $b_i$ 's which capture the local variations of intensities (Fig. 3c). The dark regions corresponds to decreasing intensity and white the ones correspond to increasing intensities. These maps can help the biologist to analyze the dynamical content of the video-microscopy image sequence. Fig. 3d shows the maps of residuals  $r_i(t)$  at time  $t = 75$ . It clearly enhances the vesicles. Each point of Fig. 3e represents the lowest non null absolute value  $|z_{k_0}|$  corresponding to the background model  $M_{k_0}$  selected at point  $p_i$  as defined in Eq (9).



**Fig. 3.** Results of the background estimation.

If  $r_i(t)$  is lower than value  $|z_{k_0}|$  at point  $p_i$ ,  $p_i$  is considered as belonging to the background. Consequently, the detected vesicles are formed by the points which do not belong to the background (Fig. 3f). The computation time on a standard recent PC (3Ghz/1Go RAM) is 6 minutes for a sequence of 150 images of size  $348 \times 260$ .

## 7. CONCLUSION

We have designed a model of image sequence for fluorescence microscopy. We have proposed an original and effective method to estimate the model parameters. It involves an asymmetric robust M-estimator and the spatial coherence of the parameters is introduced through the bias-variance trade-off principle. Reported results demonstrates that our method is able to decompose the sequence in two components: background and moving vesicles. The proposed framework is not restricted to a linear temporal evolution but can be easily extended to other evolution models (*e.g.* exponential) and finally kernel regression would be able to take into account any kind of background evolution model [7]. The decomposition of the sequence into objects and background is promising for the automatic analysis of the dynamical content of time-lapse confocal microscopy image sequences.

## 8. REFERENCES

- [1] C. Stauffer and W.E. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 747–757, Aug. 2000.
- [2] A. Mittal and N. Paragios, "Motion-based background subtraction using adaptive kernel density estimation," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, CVPR'2004*, June 2004, vol. 2, pp. 302–309.
- [3] O. Lepski, "Asymptotically minimax adaptive estimation 1: upper bounds," *SIAM Journ. Theory of Probability and Application*, vol. 36, no. 4, pp. 654–659, 1991.
- [4] C. Kervrann and J. Boulanger, "Optimal spatial adaptation for patch-based image denoising," *IEEE Trans. on Image Processing*, to appear.
- [5] S.-S. Ieng, J.-P. Tarel, and P. Charbonnier, "Evaluation of robust fitting based detection," in *Proc. of 8th Eur. Conf. on Computer Vision, ECCV'2004*, Prague, May 2004, vol. 2, pp. 341–352.
- [6] H. Allende, A. Frery, J. Galbiati, and L. Pizarro, "M-estimator with asymmetric influence function: the  $G_a^0$  distribution case," *Journ. of Statistical Computation and Simulation*, in press.
- [7] A. Ruckstuhl, M. Jacobson, R. Field, and J. Dodd, "Baseline subtraction using robust local regression estimation," *Journ. of Quantitative Spectroscopy and Radiative Transfer*, vol. 68, no. 2, 2001.
- [8] T. Gasser, L. Sroka, and C. Jennen Steinmetz, "Residual variance and residual pattern in nonlinear regression," *Biometrika*, vol. 73, pp. 625–633, 1986.
- [9] P.J. Huber, *Robust Statistics*, John Wiley and Sons, NY, 1981.
- [10] M. Maurizot, P. Bouthemy, B. Delyon, A. Iouditski, and J.M. Odobez, "Determination of singular points in 2D deformable flow fields," in *Proc. of IEEE Int. Conf. on Image Processing, ICIP'1995*, Washington DC, Oct. 1995, vol. 3, pp. 488–491.
- [11] C. Ercole, A. Foi, V. Katkovnik, and K. Egiazarian, "Spatio-temporal pointwise adaptive denoising in video: 3d non parametric approach," in *Proc. of the 1st International Workshop on Video Processing and Quality Metrics for Consumer Electronics, VPQM'2005*, Scottsdale, 2005.
- [12] S. Huet, "Model selection for estimating the non zero component of a gaussian vector," Tech. Rep. 25, University of Paris-Sud, Math. Dept., 2002.
- [13] L. Birgé and P. Massart, "Gaussian model selection," *Journ. of European Mathematical Society*, vol. 3, pp. 203–268, 2001.