TRACKING OF MULTIPLE FLUORESCENT BIOLOGICAL OBJECTS IN THREE DIMENSIONAL VIDEO MICROSCOPY

A. Genovesio[1-2], *B. Zhang*[1] *and J.-C. Olivo-Marin*[1]

1 - Unité d'Analyse d'Images Quantitative, Institut Pasteur, 25 rue du Dr Roux 75724 Paris, France
2 - SIP-CRIP5, Université René Descartes, 45 rue des Saints Pères 75270 Paris cedex 06, France

ABSTRACT

We present a method which allows for the first time to perform automaticaly the detection and the tracking of microscopic objects directly from three dimensional image data. It enables to analyse biological moving objects in three dimensional fluorescence image sequences coming from biological immunomicroscopy experiments, and get quantitative data such as the number of objects, their position, movement phases and speed.

After a detection step is performed through the multiscale analysis of images using a shift-invariant wavelet transform, the tracking is achieved using a Kalman filter and an association which enable the position of the moving objects to be predicted, refined and updated. Trajectories are analysed in terms of different parameters relevant for the motility analysis of biological objects.

1. INTRODUCTION

Presently, a considerable part of biological imaging is shifting towards *in vivo* multidimensional microscopy, allowing for the visualisation of biological processes in real time and in living systems. From life-visualisation of host-pathogen interactions down to the monitoring of protein-protein interactions, novel microscopy techniques like optical sectioning or confocal microscopy, allow for a comprehensive and effective documentation of dynamic processes by providing three dimensional spatial information and an increased spatial resolution. The arrival of these new generations of microscopes into the laboratory has opened the road to a whole series of research perspectives dedicated to the study of cellular dynamics, and of the links between cellular functions and their spatial-temporal localisation. It is therefore of utmost importance to develop image processing and analysis methods that, in order to perform the quantitative analysis of temporal image sequences in multidimensional microscopy, are able to take into account not only the spatial dimensions, but the temporal and multispectral ones as well.

A.G. was supported by IP and ANRS through an AC-14.2 fellowship.
E-mail: {agenoves,jcolivo}@pasteur.fr

Here, for the first time, we report a method which allows to perform in a fully automatic manner the detection and the tracking of microscopic objects directly from three dimensional image data. This is in contrast with previously reported methods, including ours [2], that were able to process two dimensional data only [8].

Within a biological context, identification and individualisation of moving objects is equivalent to tracking in the three dimensions several objects that are moving in different directions at different speeds and that can aggregate or disappear temporarily. To make that problem tractable, we have implemented a two steps procedure: firstly, the objects are detected in the image stacks thanks to a three-dimensional wavelet transform. In a following step, spatial positions are linked in order to compute trajectories by using a Kalman filtering step followed by an association procedure based on the minimisation of a cost function.

2. SPOT DETECTION

Spot detection is based on a multiscale approach that uses a shift invariant discrete wavelet transform(SI-DWT)[6, 7] and on the selective filtering of wavelet coefficients. This scheme allows to separate and characterize objects of different sizes by selecting only a vicinity of detail images with corresponding scales adapted to the size of the spots [6]. The extraction step consists in retaining the significant responses of the locally supported detail signal filter to the desired features, at the different scales of the wavelet representation. This is accomplished through a denoising technique using a threshold value which is image and level dependent and which can be computed automatically from the data. We consider that the input image X is contaminated by an additive gaussian noise n, $Y = X + n$, where Y is the observation and after the wavelet transform, we have $\mathbf{W}Y = \mathbf{W}X + \mathbf{W}n$ where $\mathbf{W}$ is the wavelet transform matrix. Assuming that the noise is stationary in each scale, we used Jeffreys' noninformative threshold[4] as an estimation method of the coefficients $\mathbf{W}Y_i$ at a given scale i:

$$t_{noninformative}(\mathbf{W}Y_i) = \frac{(\mathbf{W}Y_i^2 - 3\sigma_i^2)_+}{\mathbf{W}Y_i} \qquad (1)$$

where σ_i is the standard deviation of noise at scale i in the wavelet domain. A robust estimation[3] of σ_i is calculated from the median of the absolute value of the wavelet coefficients at scale i:

$$\tilde{\sigma}_i = \frac{median(|\mathbf{W}Y_i|)}{0.6745} \quad (2)$$

To characterize the spots, we finally compute a correlation image $P_s(x,y,z)$ which is the direct multiplication of s images corresponding to the selected scales:

$$P_s(x,y,z) = \prod_{i=j_1}^{j_s} \mathbf{W}Y_i(x,y,z)_+ \quad (3)$$

The correlation not only significantly reduces the remaining noise, but it also increases the contrast between the objects and the background.

3. TRACKING

Once spot positions have been computed by the detection procedure, it is necessary to link the successive spatio temporal coordinates in order to create valid trajectories. In order to do this, at each frame, we need to: 1) create a prediction of the estimation of each object in the next frame; 2) match each predicted estimation with a detected object in the next frame; and 3) update each estimation in order to built a better prediction at the next step. In the following, we describe the principles of the Kalman filter, the choice of a cinematic model well adapted to our objects and the association procedure.

The Kalman filter is a technique used to estimate values that vary over time. We now recall its principles[1, 2] and describe the way we used it in our method. In the following, the discrete time is denoted t. Vectors are written with lower case bold letters: $\mathbf{x}$, matrices are written with upper case bold letters $\mathbf{F}$ and the transpose is denoted with ^{T}. The notations $\hat{\mathbf{x}}(t+1|t)$ (vector) and $\mathbf{P}(t+1|t)$ (matrix) denote respectively the conditional expectation of $\mathbf{x}$ at time $t+1$ given $\mathbf{Z}^t$ and the conditional expectation of $\tilde{\mathbf{x}}(t+1|t)\tilde{\mathbf{x}}(t+1|t)^T$ at time $t+1$ given $\mathbf{Z}^t$ where $\mathbf{Z}^t = \{\mathbf{z}(j), j = \{1,...,t\}\}$. The estimation of the vector $\mathbf{x}$ at time t is written $\hat{\mathbf{x}}(t|t)$, the prediction of the vector $\mathbf{x}$ at time t is written $\hat{\mathbf{x}}(t+1|t)$, the error of estimation is written $\tilde{\mathbf{x}}(t|t)$ (where $\tilde{\mathbf{x}}(t|t) = \mathbf{x}(t) - \hat{\mathbf{x}}(t|t)$) and the error of prediction is written $\tilde{\mathbf{x}}(t+1|t)$ (where $\tilde{\mathbf{x}}(t+1|t) = \mathbf{x}(t+1) - \hat{\mathbf{x}}(t+1|t)$). The system state is expressed as:

$$\begin{cases} \mathbf{x}(t+1) &= \mathbf{F}(t)\mathbf{x}(t) + \nu(t) \\ \mathbf{z}(t) &= \mathbf{H}(t)\mathbf{x}(t) + \mu(t) \end{cases} \quad (4)$$

Given the current time t, equation (4) gives the new state $\mathbf{x}(t+1)$ of the system at time $t+1$ (the state $\mathbf{x}$ contains the values that describe an object; for instance position, speed, intensity, etc...), where $\mathbf{x}(t)$ is the previous state. $\mathbf{F}(t)$ is the transition matrix which models the evolution of the system and computes the next state from the previous one.
Equation (4) gives also the measurement $\mathbf{z}$, where $\mathbf{H}$ is the observation matrix. $\mathbf{H}$ is an image of the observability of the system; it is a diagonal matrix equal to matrix identity (denoted $\mathbf{I}$) if the system is fully observable.

The last term of each equation, $\nu(t)$ and $\mu(t)$, are respectively the state noise (disturbance on the system) and the measurement noise. Both noises are assumed random, uncorrelated ($E[\nu(i)\mu(j)] = 0\ \forall i$) and centered ($E[\nu(i)] = 0, E[\mu(i)] = 0\ \forall i$); their covariance is written:

$$E[\nu(i)\nu^T(j)] = \begin{cases} \mathbf{Q}(i) & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad (5)$$

$$E[\mu(i)\mu^T(j)] = \begin{cases} \mathbf{R}(i) & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad (6)$$

The algorithm is iterative and each loop has three steps: calculation of the initial predictions, selection by the association step of the best adapted measurement $\mathbf{z}$ for each object (relative to the prediction and to the other objects) and refinement of the estimations.
There are two predictions: the prediction of states and the prediction of the state error covariance matrix. They are given respectively by:

$$\hat{\mathbf{x}}(t+1|t) = E[\mathbf{x}(t+1)|\mathbf{Z}^t] \quad (7)$$

$$\mathbf{P}(t+1|t) = E[\tilde{\mathbf{x}}(t+1)\tilde{\mathbf{x}}^T(t+1)|\mathbf{Z}^t] \quad (8)$$

and can be computed recursively using:

$$\hat{\mathbf{x}}(t+1|t) = \mathbf{F}(t)\hat{\mathbf{x}}(t|t) \quad (9)$$

$$\mathbf{P}(t+1|t) = \mathbf{F}(t)\mathbf{P}(t|t)\mathbf{F}^T(t) + \mathbf{Q}(t) \quad (10)$$

where $\hat{\mathbf{x}}(t|t)$ and $\mathbf{P}(t|t)$ are the refined estimations obtained at the previous iteration.

The innovation is given by:

$$\tilde{\mathbf{z}}(t+1|t) = \mathbf{z}(t+1) - \mathbf{H}(t+1)\hat{\mathbf{x}}(t+1|t). \quad (11)$$

The covariance of the innovation $\mathbf{S}$ is expressed as:

$$\mathbf{S}(t+1) = \mathbf{H}(t+1)\mathbf{P}(t+1|t)\mathbf{H}^T(t+1) + \mathbf{R}(t+1) \quad (12)$$

This last expression is used to calculate the Kalman gain $\mathbf{K}$ that will allow the prediction to be refined as shown in equations (14) and (15):

$$\mathbf{K}(t+1) = \mathbf{P}(t+1|t)\mathbf{H}^T(t+1)\mathbf{S}^{-1}(t+1) \quad (13)$$

The refined estimation $\hat{\mathbf{x}}(t+1|t+1)$ of the prediction $\hat{\mathbf{x}}(t+1|t)$ takes into account the observation $\mathbf{z}(t+1)$; it is given by:

$$\hat{\mathbf{x}}(t+1|t+1) = \hat{\mathbf{x}}(t+1|t) + \mathbf{K}(t+1)\tilde{\mathbf{z}}(t+1|t) \quad (14)$$

The refinement $\mathbf{P}(t+1|t+1)$ of the covariance matrix of the errors of the state $\mathbf{P}(t+1|t)$ is:

$$\mathbf{P}(t+1|t+1) = [\mathbf{I} - \mathbf{K}(t+1)\mathbf{H}(t+1)]\mathbf{P}(t+1|t) \quad (15)$$

Equation (14) shows that the refinement of the state prediction is the addition of the first prediction (which trusts the model) and of the innovation (assessment of the prediction error) weighted by the gain matrix $\mathbf{K}$. Depending on the quality of the model describing the data, the gain matrix will allow (or not) the refined estimation to get closer to the first prediction.

A state $\mathbf{x}(t)$ (respectively a measurement $\mathbf{z}(t)$) is a vector of R^n where each coordinate represents a characteristic (respectively a measure of a characteristic) of an object at time t. Typically, the three first coordinates represent the 3D-space location. In our application, two dimensions are added to the measurement vector to represent the area (num. of pixels) and the intensity (mean intensity on the area) of a spot as follows:

$$\mathbf{z}(t)^T = [x_t, y_t, z_t, a_t, i_t] \quad (16)$$

A model $\mathcal{M} = (\mathbf{F}, \mathbf{H}, \mathbf{Q}, \mathbf{R})$ is a set of matrices that represent the dynamics, the observation and the noises of the process under estimation. For example, one may consider the following expression for $\mathbf{F}$:

$$\mathbf{x}(t) = \begin{bmatrix} x_t \\ y_t \\ z_t \\ x_{t-1} \\ y_{t-1} \\ z_{t-1} \end{bmatrix}, \mathbf{F} = \begin{bmatrix} 2 & 0 & 0 & -1 & 0 & 0 \\ 0 & 2 & 0 & 0 & -1 & 0 \\ 0 & 0 & 2 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \quad (17)$$

which represents a linear extrapolation of the last two points. When there is no *a priori* information about the movement of the objects in a sequence, the transition matrix is best represented by the identity matrix. This means that the best prediction that can be obtained from the computation is the last estimation itself and that the movement is random.

The challenge therefore is to find a model $\mathcal{M}$ that best matches the dynamics of the process. This is accomplished by finding isolated objects which are used to test several type of predefined models. The benefit of using isolated objects is that the detections do represent the same object without ambiguities and therefore that the computed tracks are correct. To compare models and select the one which is best adapted, we compute the first two moments of the error between the Kalman filter prediction and the real measurements. The model which minimizes these values is selected.

The association stage consists of finding the best matches between the detections and the predictions obtained with the Kalman filter. The best match, however, does not mean that an optimal global matching, like Jonker and Volgenant's algorithm [5] produces, is achieved. This is due to the fact that generally the number of predictions never equals the number of measurements. Indeed, objects may leave the volume and as a consequence the measurements for the tracks associated to them would not be available. Likewise, if an object enters the volume, a new track is to be initiated without previous measurements associated to it. Also, if a detection originates from noise or spurious data, it should not be associated with a track. Accordingly, the assignment procedure is as follows: first, we compute all the costs between measurements $\{\mathbf{z}_j(t+1), j \in \{1, \cdots, n\}\}$ and predicted measurements $\{\hat{\mathbf{z}}_i(t+1|t), i \in \{1, \cdots, m\}\}$, given by:

$$\mathbf{C} = \begin{bmatrix} & \vdots & \\ \cdots & \text{cost}(\hat{\mathbf{z}}_i(t+1|t), \mathbf{z}_j(t+1)) & \cdots \\ & \vdots & \end{bmatrix} \quad (18)$$

where cost($\mathbf{a}$, $\mathbf{b}$) is a generalised distance function, with the Euclidian distance being used for the spatial coordinates. The assignments $(a_0, ..., a_l)$ (where $l = \min(n, m)$) are then found according to:

$$a_k = (i_k, j_k) = \arg \min_{(i,j) \in \mathbf{J}_k} \mathbf{C}(i, j) \quad (19)$$

with

$$\mathbf{J}_0 = \{(i,j)/i \in \{1, \cdots, m\}, j \in \{1, \cdots, n\}\} \quad (20)$$

$$\mathbf{J}_{k+1} = \mathbf{J}_k \backslash \{(i,j)/\{i = i_k\} \cup \{j = j_k\}\} \quad (21)$$

The search for assignments stops when either $\mathbf{J}_k = \{\emptyset\}$ or $\mathbf{C}(i_k, j_k) > \eta$, with η a fixed threshold which represents the maximum cost between a detection and its prediction. This assignment technique is a sub optimal one in a global sense, that is to say, it may occur that

$$\exists \{b_0, .., b_{\min(n,m)}\}, \sum_k \mathbf{C}(b_k) < \sum_k \mathbf{C}(a_k) \quad (22)$$

but it does produce however much better results than an optimal assignment.

4. RESULTS

In order to assess the quality of our method, we have generated sequences with artificial objects whose characteristics are as close as possible to real biological objects. The objects are represented by 3D Gaussian shapes with different random covariance matrices in order to get different shapes of different sizes (from 8x8x8 to 20x20x20 voxels) and are included in a 100x100x15 volumetric image. Several sequences of 50 volumetric images were generated where the

objects were made to move randomly following two different models of movement, with an additionnal random modification of the covariance matrix with time to make their aspect change. Also, white Gaussian noise $\mathcal{N}(128, 10)$ was added to the sequence in order to represent the noise present in typical microscopy images. Figure 1 *top* shows different views of a frame, where the high density of objects and the high level of noise can be appreciated. Figure 1 *bottom* shows the quality of object detection achieved with our method. Figure 2 show the trajectories obtained when performing the tracking on the whole sequence. Finally, Figure 3 gives a summary of the influence of the number of objects within the same volume on the performances of the algorithm. The results are shown for 10, 20, 40 and 60 objects, and it can be appreciated that even at a high density of objects the performances of the algorithm are correct.

5. CONCLUSION

We have presented a method to detect and track microscopic objetcs directly from three dimensional image data in biological immunomicroscopy images. The method uses a shift-invariant wavelet transform for the detection and a Kalman filtering technique associated to the minimization of a cost function for the tracking. We have shown on generated sequences that thanks to a local assigment association algorithm, it is possible to establish valid trajectories even in the cases where the high density of objects gives ambiguous detection measurements.

6. REFERENCES

[1] Y. Bar-Shalom and T.E. Fortmann. *Tracking and Data Association*. Academic Press, New-York, 1988.

[2] Briquet-Laugier, F., Boulin C., and Olivo-Marin J.-C. Analysis of moving biological objects in video microscopy sequences. *Proc. of SPIE, Vol. 3642*, pages 1–11, 1998.

[3] D.L. Donoho and I.M. Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994.

[4] M.A.T. Figueiredo and R.D. Nowak. Wavelet-based image estimation: An empirical bayes approach using jeffrey's noninformative prior. *IEEE Trans. on Image Processing*, 10(9):1322–1331, 2001.

[5] R. Jonker and A. Volgenant. A shortest augmenting path algorithm for dense and sparse linear assignment problems. *Computing*, 38:325–340, 1987.

[6] J.-C. Olivo-Marin. Extraction of spots in biological images using multiscale products. *Pattern Recognition*, 112(6):1–2, 2002.

[7] J.-L. Starck, F. Murtagh, and A. Bijaoui. *Image Processing and Data Analysis: the Multiscale Approach*. Cambridge University Press, 2000.

[8] W. Tvarusko, M. Bentele, T. Misteli, R. Rudolf, C. Kaether, D. L. Spector, H. H. Gerdes, and R. Eils. Time-resolved analysis and visualization of dynamic process in living cells. *PNAS*, 96:7950–7955, 1999.

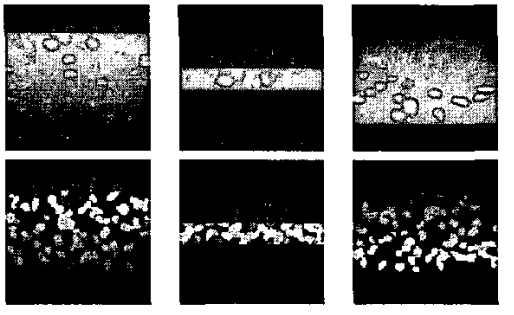

Fig. 1. Object detection in 3D stack. Top: original noisy frame. Bottom: detected objects. On both: angle of view -45,0,45

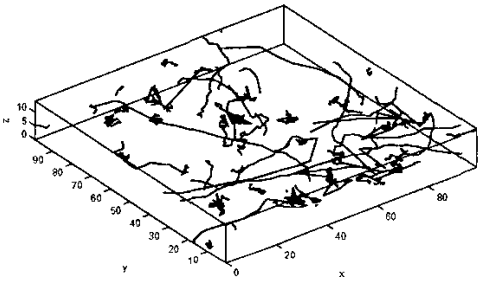

Fig. 2. Tracks of objects in a 3D sequence.

Nb of obj.	% correctly tracked
10	95%
20	87%
40	71%
60	59%

Fig. 3. Influence of the number of objects on the quality of tracking, in a 100x100x15 voxels volume.